

METHODISCHE FRAGEN BEI DER ERFASSUNG VON EMOTIONALER INTELLIGENZ

Dissertation zur Erlangung des Doktorgrades
der Naturwissenschaften (Dr. rer. nat.)

dem Fachbereich Psychologie der Philipps-Universität Marburg
vorgelegt von

Diplom-Psychologin Tonja Berit Plücken
geboren in Tübingen

Stuttgart, im Januar 2015

Vom Fachbereich Psychologie der Philipps-Universität Marburg als Dissertation am 13. April 2015
angenommen.

Erstgutachter: Prof. Dr. Lothar Schmidt-Atzert, Philipps-Universität Marburg

Zweitgutachterin: Prof. Dr. Ricarda Steinmayr, Technische Universität Dortmund

Tag der mündlichen Prüfung: 13. April 2015

Vorwort

„Die Neugier steht immer an erster Stelle eines Problems, das gelöst werden will“ (Galileo Galilei). So stand die Neugier darüber, was sich hinter dem Konstrukt Emotionale Intelligenz verbirgt, an erster Stelle dieser Dissertation. Ausgelöst wurde die Neugier vor allem dadurch, dass sich „Emotionale Intelligenz“ immer mehr in den populärwissenschaftlichen Medien und in der allgemeinen Öffentlichkeit verbreitete, das Konstrukt im Rahmen des Psychologiestudiums in den Vorlesungen und Seminaren jedoch nur marginal thematisiert wurde. Die vorliegende Dissertation beschäftigt sich daher mit diesem noch immer vergleichsweise neuen und vor allem populären Konstrukt. Obwohl in den letzten Jahren viele Bemühungen zur Erforschung von Emotionaler Intelligenz zu verzeichnen sind, gibt es nach wie vor diverse offene Fragen zum Konstrukt an sich und zu dessen Erfassung (Maul, 2011). Roberts, Zeidner und Matthews (Roberts, Zeidner, & Matthews, 2001) gehen jedoch davon aus, dass wesentliche Kritikpunkte am Konstrukt in erster Linie auf technisch-methodische Probleme zurückgehen, die mit Fortschritten in der Testentwicklung behoben werden können. Es sei daher wünschenswert, wenn sich diese Fortschritte nicht nur auf die Optimierung existierender Verfahren beschränken, sondern auch „echte“ Alternativen zu den wenigen bestehenden Instrumenten entwickelt würden. Davies, Stankov und Roberts merken hierzu an, dass es für die Analyse des psychometrischen Status von Emotionaler Intelligenz notwendig sei „... to (a) have an adequate methodology for this construct’s assesement and (b) demonstrate it partial (or complete) independence from other [...] concepts.“ (1998, p. 989).

Zum Zeitpunkt der ersten Literaturrecherche und Auseinandersetzung mit Emotionaler Intelligenz im Rahmen des Promotionsvorhabens basierten fast alle Erkenntnisse zu diesem Konstrukt auf Studien, in denen der Test MSCEIT (Mayer-Salovey-Caruso Emotional Intelligence Test; Mayer, Salovey, & Caruso, 2002) bzw. das Vorgängerinstrument MEIS (Multifactor Emotional Intelligence Scale; Mayer, Caruso, & Salovey, 2000) zum Einsatz kamen. Daraufhin fiel die Entscheidung, dass ein Ziel dieser Dissertation die Entwicklung eines Leistungstests zur Erfassung einer Komponente Emotionaler Intelligenz darstellen sollte, um mit einem neuen Verfahren die Validierung des Konstrukts voranzutreiben. Die Zielsetzung der Arbeit wurde jedoch verändert, nachdem im Laufe der Testentwicklung verschiedene methodische Unklarheiten und Forschungsfragen auftauchten. Auf dieser Basis erschien es sinnvoll und erforderlich, den Fokus der vorliegenden Arbeit zu verändern: die Testentwicklung wird in Kapitel 1.5 beschrieben, den Kern der Arbeit bilden jedoch methodische Fragestellungen, die im Rahmen der Testentwicklung bedeutsam sind und für die es in der Literatur zu Emotionaler Intelligenz noch keine hinreichenden bzw. eindeutigen Antworten gibt. Dementsprechend wurde auch der Titel der Arbeit „Methodische Fragen bei der Erfassung von

Emotionaler Intelligenz“ gewählt. Anhand des neu entwickelten Leistungstests ViTEIP (Videobasierter Test zur Erfassung der Emotionalen Intelligenz im Pflegeberuf) und einem weiteren Verfahren zur Erfassung von Emotionaler Intelligenz (TEMINT; Schmidt-Atzert & Bühner, 2002) werden methodischen Fragen zu Auswertungsstrategien, Antwortformaten und Präsentationsmodi untersucht.

Die Forschung zu Emotionaler Intelligenz hat sich in den letzten Jahren kontinuierlich weiterentwickelt, so dass mittlerweile deutlich mehr Studien und Ergebnisse existieren. Würde man die vorliegende Arbeit mit Kenntnis der Ergebnisse dieser Studien heute konzipieren, fielen möglicherweise die Versuchsplanung oder Testgestaltung an der einen oder anderen Stelle etwas anders aus. Die Fragestellungen zu den grundlegenden methodischen Fokusthemen sind – wie nachfolgend zu lesen sein wird – jedoch nach wie vor aktuell. Ein Zitat, das zum Abschluss dieses Vorworts den Status von Emotionaler Intelligenz treffend beschreibt, findet sich bei Joseph und Newman: „... *Consequently, the current state of EI¹ is somewhat paradoxical; although EI is a wildly popular tool in organizations, organizational science has yet to answer many theoretical, measurement, and validity questions surrounding the construct...*“ (Joseph & Newman, 2010, p. 54).

¹ EI= emotional intelligence (Anmerkung der Autorin)

Danksagung

Auf dem langen Weg bis zur Vollendung meiner Dissertation wurde ich durch viele Menschen unterstützt. Ihnen allen gilt mein großer Dank.

Einigen Personen möchte ich an dieser Stelle namentlich danken: Herzlichen Dank für Rat und Tat, Geduld und Verständnis, Unterstützung und Aufmunterung! Ihr habt alle auf eure jeweils individuelle Art und Weise dazu beigetragen, mir die Fertigstellung meiner Dissertation zu ermöglichen.

Prof. Dr. Lothar Schmidt-Atzert

Prof. Dr. Ricarda Steinmayr

Pamela Holtus, Friederike Thomsen und Dajana Henkel

alle freiwilligen Teilnehmer meiner Untersuchungen

Reinhild Graser und Josef Sturm

Dr. Udo Plücken und Regina Dorn

Alina Plücken

Selina Bauer und Julia Otto

Theo Graser

und in ganz besonderem Maße: Markus Keck

HERZLICHEN DANK

Inhaltsverzeichnis

Vorwort	i
Danksagung	iii
Inhaltsverzeichnis.....	iv
Tabellenverzeichnis	vii
Abbildungsverzeichnis	ix
Abkürzungs- und Testverzeichnis	x
Zusammenfassung.....	xii
Summary	xiv
1 „What is this thing called emotional intelligence?“ – Theoretischer Hintergrund der Arbeit.....	1
1.1 Definition und Modelle von Emotionaler Intelligenz.....	1
1.2 Die Messung von Emotionaler Intelligenz	4
1.2.1 Auswertungsstrategien von Emotionale-Intelligenz-Tests	5
1.3 Beispiele für Leistungstests zur Erfassung von Emotionaler Intelligenz.....	10
1.3.1 Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT)	11
1.3.2 Test zur Erfassung der Emotionalen Intelligenz (TEMINT)	12
1.3.3 Situational Test of Emotion Management (STEM) und Situational Test of Emotional Understanding (STEU)	14
1.4 Die Validität von Emotionale-Intelligenz-Tests	16
1.5 Der ViTEIP als neuer Leistungstest zur Erfassung von Emotionaler Intelligenz	25
1.5.1 Theoretische Basis und a priori Annahmen zur Testkonstruktion	25
1.5.2 Testentwicklung ViTEIP-V1	28
1.5.3 Endversion ViTEIP-V2	31
2 Forschungsfragen und Hypothesen	34
2.1 Auswertungsstrategien	35
2.1.1 Entwicklung von Scorings mit biasfreien Kennwerten	38
2.1.1.1 Profilauswertung	39
2.1.1.2 Übertragung der Signalentdeckungstheorie auf Emotionale-Intelligenz-Tests	40
2.2 Antwortformate	42
2.3 Präsentationsmodus	43
3 Methoden	45
3.1 Stichprobenbeschreibung und Versuchsablauf	45

3.2	Testauswahl und –beschreibung	47
3.3	Vorgehen bei der Datenauswertung.....	51
3.3.1	Auswertung der Emotionale-Intelligenz-Leistungstests.....	51
3.3.1.1	Standardisierung der Daten für den Vergleich der Auswertungsstrategien.....	54
3.3.2	Auswertungen der weiteren eingesetzten Verfahren.....	55
4	Ergebnisse	56
4.1	Vergleich der Auswertungsstrategien.....	56
4.1.1	Deskriptive Statistiken und Exploration der faktoriellen Struktur	56
4.1.2	Korrelationsanalysen.....	59
4.2	Entwicklung von Scorings mit biasfreien Kennwerten.....	62
4.2.1	Profilauswertung	65
4.2.1.1	Deskriptive Statistiken.....	65
4.2.1.2	Überprüfung der Kennwerte in Bezug auf den Antwortbias	66
4.2.1.3	Korrelationsanalysen.....	67
4.2.2	Übertragung der Signalentdeckungstheorie auf Emotionale-Intelligenz-Tests	68
4.2.2.1	Deskriptive Statistiken.....	70
4.2.2.2	Detektionsleistung, Emotionale Sensitivität und Emotionale Spezifität als Leistungs- indikatoren	71
4.3	Antwortformate	75
4.4	Vergleich der Präsentationsmodi.....	78
5	Diskussion	81
5.1	Auswertungsstrategien	82
5.1.1	Vier target- und konsensbasierte Scorings im Vergleich.....	82
5.1.2	Zwei neue Scorings mit biasfreien Kennwerten.....	90
5.1.2.1	Profilauswertung	91
5.1.2.2	Übertragung der Signalentdeckungstheorie auf Emotionale-Intelligenz-Tests	93
5.1.2.3	Die Kennwerte Emotionale Sensitivität und Emotionale Spezifität	96
5.1.3	Fazit: Auswertungsstrategien.....	101
5.2	Antwortformate	102
5.2.1	Fazit: Antwortformate.....	105
5.3	Vergleich der Präsentationsmodi.....	106
5.3.1	Fazit: Präsentationsmodus	109
5.4	Kritische Diskussion der Validität von ViTEIP-V2 und TEMINT	110
5.5	Anregungen für zukünftige Studien	116

5.6 Limitationen	118
6 Anhang.....	121
7 Literaturverzeichnis	158
8 Erklärung.....	172
9 Lebenslauf Tonja Berit Plücken	173

Tabellenverzeichnis

Tabelle 1: Tabellarische Übersicht über Korrelationen der zitierten Studien.....	23
Tabelle 2: Stichprobenbeschreibung der Originalpersonen	29
Tabelle 3: Tabellarischer Überblick über die Hypothesen der vorliegenden Arbeit	35
Tabelle 4: Übertragung des Vierfelder-Schemas einer Entdeckungsaufgabe auf die Systematik der Emotionale-Intelligenz-Tests bei Targetscoring.....	40
Tabelle 5: Stichprobenbeschreibung.....	46
Tabelle 6: Eingesetzte Testverfahren in den Untersuchungen 1 und 2.	48
Tabelle 7: Deskriptive Statistiken ViTEIP-V2a und TEMINT bei verschiedenen Scorings	57
Tabelle 8: Korrelation von ViTEIP-V2 und TEMINT bei unterschiedlichen Auswertungsmethoden	60
Tabelle 9: Korrelation des ViTEIP-V2a und TEMINT (unterschiedliche Scorings) mit anderen Verfahren bzw. demographischen Variablen.....	61
Tabelle 10: Korrelationen, die sich bei verschiedenen Scorings signifikant unterscheiden	62
Tabelle 11: Korrelation der Kennwerte zum Antwortverhalten mit den Gesamttestwerten bei unterschiedlichen Scorings (ViTEIP-V2 bzw. TEMINT)	64
Tabelle 12: Korrelation der Kennwerte zum Antwortverhalten (ViTEIP-V2 und TEMINT, Stichprobe 1)	64
Tabelle 13: Deskriptive Statistiken sowie Mittelwertvergleiche zwischen den Stichproben bei Profilauswertung (ViTEIP-V2 und TEMINT)	65
Tabelle 14: Korrelation „Profil-Gesamtwert“ mit Kennwerten zum Antwortverhalten von ViTEIP-V2 und TEMINT	66
Tabelle 15: Korrelationen Target-Übereinstimmung und Profil-Gesamtwert (ViTEIP-V2 und TEMINT)	67
Tabelle 16: Kennwerte aus der Signalentdeckungstheorie in der Übertragung auf Emotionale- Intelligenz-Tests.....	68
Tabelle 17: Kennwerte der Signalentdeckungstheorie in der Übertragung auf ViTEIP-V2 und TEMINT	71
Tabelle 18: Korrelation der Kennwerte zum Antwortverhalten zu „Detektionsleistung“ (ViTEIP- V2 und TEMINT)	72
Tabelle 19: Korrelation der Kennwerte zum Antwortverhalten zu „Emotionale Sensitivität“ bzw. „Emotionale Spezifität“ (ViTEIP-V2 und TEMINT)	73
Tabelle 20: Korrelation der Kennwerte ViTEIP-V2 und TEMINT mit anderen Verfahren (Stichprobe 1)	74

Tabelle 21: Deskriptive Statistiken des ViTEIP-V2 mit den Antwortformaten „Ankreuzen“ und „Zuordnen“ bei unterschiedlichen Scorings.....	76
Tabelle 22: Korrelationen der video- und textbasierten Version des ViTEIP mit anderen Kennwerten sowie Prüfung auf Unterschiedlichkeit der Korrelationen	80
Tabelle 23: Tabellarischer Überblick über die Hypothesen und deren Beurteilung	81

Abbildungsverzeichnis

Abbildung 1: Vereinfachte Darstellung des Fähigkeitsmodells der Emotionalen Intelligenz nach Mayer & Salovey (1997).....	3
Abbildung 2: Schematische Darstellung der unterschiedlichen Forschungsrichtungen und Umsetzung in Messinstrumente zu Emotionaler Intelligenz	4
Abbildung 3: Übersicht über den MSCEIT V2.0 inkl. Angaben zur Reliabilität nach Mayer et al. (Mayer J. D., Salovey, Caruso, & Sitarenios, 2003)	12
Abbildung 4: Beispielsituation des TEMINT inkl. Auswertung Target-Abweichung.....	13
Abbildung 5: Beispielsituation des STEM (siehe: Ferguson & Austin, 2011).....	14
Abbildung 6: Beispielsituationen des STEU für verschiedene Kontexte neutral/privat/beruflich (MacCann & Roberts, 2008).....	15
Abbildung 7: Ablauf der Videosequenzen am Beispiel der Situation 7 „Frau Hoffmann bestellt Inkontinenzeinlagen“	30
Abbildung 8: Schematische Darstellung der Rausch- und Signalverteilung bei einem Signal-entdeckungsexperiment in der Übertragung auf Emotionale-Intelligenz-Tests und verschiedenen Ausprägungen der Lage von c	41
Abbildung 9: Darstellung der Auswertung mittels Target-Übereinstimmung und Target-Treffer an einem fiktiven Beispiel.....	53
Abbildung 10: Schematische Darstellung der Rausch- und der Signalkurve für TEMINT und ViTEIP	70

Abkürzungs- und Testverzeichnis

AdO	Angabe der Originalperson	IT-Face	Inspection time faces (Inspektionszeit Gesichter) (Austin, 2004)
AVEI	Audiovisual Test of Emotional Intelligence (Zysberg, Levy, & Zisberg, 2011)	K-A	Scoring „Konsens-Anteil“
d2-R	Aufmerksamkeits-Konzentrations-Test d2, revidierte Form (Brickenkamp, Schmidt-Atzert, & Liepmann, 2010)	K-M	Scoring „Konsens-Modal“
DANVA	Diagnostic Analysis of nonverbal Accuracy (Nowicki & Duke, 1994)	LEAS	Level of Emotional Awareness Scale (Lane, Quinlan, Schwartz, Walker, & Zeitlin, 1990)
EARS	Emotional Accuracy Research Scale (Mayer & Geher, 1996)	MEIS	Multifactor Emotional Intelligence Scale (Mayer, Caruso, & Salovey, 2000)
ECI	Emotional Competence Inventory (Boyatzis, Goleman, & Rhee, 2000)	MERT	Multimodal Emotion Recognition Test (Bänziger, Grandjean, & Scherer, 2009)
EI	Emotionale Intelligenz oder emotional intelligence	MSCEIT	Mayer-Salovey-Caruso Emotional Intelligence Test (Mayer, Salovey, & Caruso, 2002)
EKF-Fremd	Emotionale-Kompetenz-Fragebogen, Fremdeinschätzung (Rindermann, 2007)	MSFDE	Montreal Set of Facial Displays of Emotion (Beaupré, Cheung, & Hess, 2000)
EKF-Selbst	Emotionale-Kompetenz-Fragebogen, Selbsteinschätzung (Rindermann, 2007)	MTMM	Multitrait-Multimethod-Analyse
EQ-360	The Bar-On EQ-360 (Bar-On & Handley, 2003)	NEO-FFI	NEO-Fünf-Faktoren Inventar nach Costa und McCrae (Borkenau & Ostendorf, 1993)
EQ-i	Emotional Quotient Inventory (Bar-On, 1997)	NEO-PI-R	Revised NEO Personality Inventory (Costa & McCrae, 1992)
eSens	Kennwert „Emotionale Sensitivität“	OCEANIC-20	Openness conscientiousness extraversion agreeableness neuroticism index condensed (Roberts R. D., 2000)
eSpez	Kennwert „Emotionale Spezifität“	OREIS	Other-report EI-scale (Choi & Kluemper, 2012)
Gf/Gc QTB	Gf/Gc Quickie Test Battery (Stankov, 1997)	SI	Soziale Intelligenz
GMA	General mental ability	SJT	Situational Judgment Test
IPT	Interpersonal Perception Task-15 (Costanzo & Archer, 1993)	SPTB	Sjöberg Personality Test Battery (Sjöberg, 2001)
IPIP	International Personality Item Pool (Goldberg, et al., 2006)	SREIS	Self-report Emotional Intelligence Scale (Brackett, Rivers, Lerner, Salovey, & Shiffman, 2006)
IST-2000-R	Intelligenz-Struktur-Test 2000-R (Amthauer, Brocke, Liepmann, & Beauducel, 2001)		

SSRI	Schutte Self-Report Inventory (Schutte, et al., 1998)	ViTEIP-V1	Videobasierter Test zur Erfassung der Emotionalen Intelligenz im Pflegeberuf; Version 1 mit 18 Situationen und Antwortformat „Ankreuzen“
STEM	Situational Test of Emotion Management (MacCann & Roberts, 2008)	ViTEIP-V2	Videobasierter Test zur Erfassung der Emotionalen Intelligenz im Pflegeberuf; Version 2 mit 12 Situationen
STEU	Situational Test of Emotional Understanding (MacCann & Roberts, 2008)	ViTEIP-V2a	Videobasierter Test zur Erfassung der Emotionalen Intelligenz im Pflegeberuf; Version 2 mit 12 Situationen und Antwortformat „Ankreuzen“
TAS-20	Toronto Alexithymia Scale (Bagby, Parker, & Taylor, 1994)	ViTEIP-V2b	Videobasierter Test zur Erfassung der Emotionalen Intelligenz im Pflegeberuf; Version 2 mit 12 Situationen und Antwortformat „Zuordnen“
TEII	Tapia Emotional Intelligence Inventory (Tapia, 2001)	WIT-2	Der Wilde-Intelligenztest (Kersting, Althoff, & Jäger, 2008)
TEIQue	Trait Emotional Intelligence Questionnaire (Petrides & Furnham, 2003)	WLEIS	Wong and Law Emotional Intelligence Scale (Wong & Law, 2002)
TEMINT	Test zur Erfassung von Emotionaler Intelligenz (Schmidt-Atzert & Bühner, 2002)	WPT	Wonderlic Personnel Test (Wonderlic Personnel Test Inc., 2000)
TMMS	Trait Meta-Mood Scale (Salovey P. , Mayer, Goldman, Turvey, & Palfai, 1995)		
T-T	Scoring „Target-Treffer“		
T-Ü	Scoring „Target-Übereinstimmung“		
ViTEIP	Videobasierter Test zur Erfassung der Emotionalen Intelligenz im Pflegeberuf; ohne spezifische Bezeichnung der Version		

Zusammenfassung

Über das Konstrukt Emotionale Intelligenz, verstanden als die Fähigkeit emotionale Informationen wahrzunehmen, zu nutzen, zu verstehen und zu regulieren (Mayer & Salovey, 1997; Mayer, Caruso, & Salovey, 2000), wird in der psychologischen Forschungsgemeinde seit nunmehr gut 25 Jahren rege und kontrovers diskutiert. Trotz großer Forschungsbemühungen können nach wie vor verschiedene Fragen zu Konzeption, Messung und Validierung nicht eindeutig oder nur unzureichend beantwortet werden. Die Forschungsfragen der vorliegenden Dissertation lauten daher: 1. Was passiert, wenn ein Emotionale-Intelligenz-Test mit unterschiedlichen Scorings ausgewertet wird und welche Scorings sind für die Anwendung zu empfehlen? 2. Inwiefern wirkt sich ein geändertes Antwortformat („Ankreuzen“ vs. „Zuordnen“) auf Ergebnisse eines Emotionale-Intelligenz-Tests, dem Videobasierten Test zur Erfassung der Emotionalen Intelligenz im Pflegeberuf (ViTEIP-V2), aus? 3. Ist ein videobasierter Präsentationsmodus für den ViTEIP-V2 geeignet?

In die erste Untersuchung flossen die Daten von $N=113$ überwiegend studentischen Teilnehmern ein. An der zweiten Untersuchung beteiligten sich insgesamt sieben Altenhilfe- und Krankenpflegeschulen, woraus sich zwei Stichproben mit $N=138$ und $N=122$ generierten. Alle Teilnehmer bearbeiteten die beiden Leistungstests zur Erfassung von Emotionaler Intelligenz ViTEIP-V2 und TEMINT (Schmidt-Atzert & Bühner, 2002) sowie weitere Verfahren.

Aus der vorliegenden Arbeit ergibt sich die Empfehlung, auf sogenannte targetbasierte Scorings zurückzugreifen, bei denen die Auswertung über einen Vergleich zwischen der Antwort der Probanden und den situationsgebenden Originalpersonen (Target) erfolgt. Neben grundlegenden kritischen Einwänden bezüglich des Konsensscorings (d.h. Verwendung der Mehrheitsmeinung zur Bestimmung der korrekten Antwort), sprechen eine höhere Varianz und Reliabilität sowie etwas deutlichere Validitätsbelege für den Einsatz von Targetscorings. Die Auswirkungen, die sich durch auf Seiten der Probanden auftretende Antworttendenzen (Bias) ergaben, konnten durch drei neu entwickelte targetbasierte Scoringprozeduren nivelliert werden.

Mit der Auswertung über Antwortprofile im Kontext der Profilauswertung wurde der Einfluss des Bias auf das Testergebnisses deutlich reduziert; aufgrund geringer Reliabilitäten ist ein Einsatz dieses Scorings nur bedingt zu empfehlen. Der für den zweiten Scoringansatz aus der Signalentdeckungstheorie entlehene Kennwert „Detektionsleistung“ ermöglichte eine überwiegend biasfreie Leistungserfassung. Für die Auswertung des ViTEIP-V2 und TEMINT zu präferieren ist aus Sicht der Autorin jedoch ein Scoring, bei dem eine Aufteilung der Items zu den zwei Kennwerten „Emotionale Sensitivität“ (Umgang mit vorhandenen Emotionen) und „Emotionale Spezifität“ (Umgang mit nicht vorhandenen Emotionen) erfolgte. Mit dieser Auswertung wurden die Effekte,

die sich aus den spezifischen Ankreuzverhalten der Probanden ergaben, deutlich minimiert; die generellen Korrelationsmuster zur Validität konnten zudem mittels der Aufteilung in zwei Kennwerte besser erklärt werden als mit nur einem Kennwert. Einige Fragen bezüglich der Validität von ViTEIP-V2 und TEMINT blieben, unabhängig vom eingesetzten Scoring, offen.

Eine Veränderung des Antwortformates beeinflusste die Probanden deutlich in ihrem Antwortverhalten. Der hohe Aufforderungscharakter bezüglich der Nutzung der Antwortkategorie „Emotion nicht vorhanden“, den das Antwortformat „Zuordnen“ im Vergleich zu „Ankreuzen“ auf die Probanden hatte, verzerrte die Abbildung der tatsächlichen Leistungsfähigkeit. Von einer Verwendung des Formates „Zuordnen“ wird daher abgeraten. Die Vorteile, die verschiedene Autoren im Einsatz von videobasierten im Vergleich zu textbasierten Tests sehen, bestätigten sich in dieser Arbeit für den ViTEIP nicht. Aus Sicht der Autorin gilt es daher sorgfältig abzuwägen, ob sich der Aufwand lohnt, den die Entwicklung und der Einsatz von videobasierten Verfahren erfordern, wenn sich keine Vorteile für diesen Präsentationsmodus ergeben.

Die vorliegende Arbeit trägt durch die Untersuchung methodischer Fragen und der Entwicklung neuer Scorings dazu bei, die Erfassung des Konstrukts Emotionale Intelligenz zu verbessern. Die Ergebnisse bezüglich der Unterscheidung in „Emotionale Sensitivität“ und „Emotionale Spezifität“ stellen die Annahme, bei Emotionaler Intelligenz handle es sich um ein eindimensionales und homogenes Konstrukt, in Frage und regen zu weiteren Forschungen an. Da nicht auszuschließen ist, dass die methodischen Aspekte Scoring, Antwortformat und Präsentationsmodus in einem komplexeren Zusammenhang stehen als aus den vorliegenden Datensätzen berechnet werden konnte, bietet es sich an, zukünftige Untersuchungen so zu konzipieren, dass alle drei Faktoren systematisch innerhalb einer Stichprobe analysiert werden können.

Summary

The construct of emotional intelligence, defined as the ability to perceive, to use, to understand and to regulate emotional information (Mayer & Salovey, 1997; Mayer, Caruso, & Salovey, 2000), has now been lively and controversially discussed for more than 25 years. Despite extensive research, different questions about conceptualization, measurement and validation can only be inconclusively or insufficiently answered. Research questions of this doctoral thesis are therefore: 1. What happens when a test of emotional intelligence is evaluated with different scorings? Which scoring is to be recommended? 2. How does a different response format (“Ticking” vs. “Allocating”) influence results of an emotional-intelligence-test, the ViTEIP-V2 (Videobasierter Test zur Erfassung der Emotionalen Intelligenz im Pflegeberuf)? 3. Is a video-based presentation appropriate for the ViTEIP-V2?

In the first study, data of $N=113$ mostly student participants were collected. In the second study, two samples of $N=138$ and $N=122$ were recruited from seven nurse's training schools. All participants completed the two performance tests of emotional intelligence ViTEIP-V2 and TEMINT (Schmidt-Atzert & Bühner, 2002) and further tests.

After data evaluation, the author recommends utilization of so-called target-based scorings. Target-based scorings involve a comparison of the participants' answers and the answer of the target-person who provided information on the situation. Beyond fundamental critical objections referring consensus-scorings (using majority view to generate the correct response-option), higher variance and reliability as well as slightly better validity favor the use of target-based scorings. Three newly developed target-based scorings were able to reduce influence of an existing participant's response bias.

Use of response profiles as scoring method significantly reduced bias. Due to low reliability of this scoring, the profile evaluation is only conditionally recommended. The parameter “Discriminability index”, deriving from signal detection theory, was used within the scope of a second scoring. It enabled a predominantly bias-free data collection. Splitting items into two parameters “Emotional Sensitivity” (dealing with existing emotions) and “Emotional Specificity” (dealing with absent emotions) led to the preferred scoring technique for ViTEIP-V2 and TEMINT. Effects on participants' answering strategies were minimized and interpretation of correlations concerning validity was much better with two parameters compared to one single parameter. Independent of the scorings, some questions regarding validity of ViTEIP-V2 and TEMINT remain unanswered.

Change of the response format from “Ticking” to “Allocating” had an influence on subjects’ responses. A high demanding character of “Allocating” compared to “Ticking” led to an increased use of the answering option “no emotion”. As actual performance capabilities were distorted, the usage of “Allocating” is not recommended. In this study, expected benefits of a video-based version compared to a text-based version of ViTEIP were not supported. If there are no advantages of a video-based test, the author suggests careful consideration of the efforts that the development and administration of such tests require.

By analyzing methodical questions and developing new scorings, the present work contributes to improve measurement of emotional intelligence. Particularly regarding results on the differentiation between “Emotional Sensitivity” and “Emotional Specificity” challenges the assumption of an unidimensional and homogenous construct of emotional intelligence. Further research is suggested, as existing data does not allow systematic analysis of scoring, response format and presentation mode within one sample. To generate future studies in a way where a possible complex relationship of all three factors can be simultaneously analyzed is recommended.

1 „What is this thing called emotional intelligence?“² – Theoretischer Hintergrund der Arbeit

Über Emotionale Intelligenz wird in der psychologischen Forschungsgemeinde seit nunmehr gut 25 Jahren rege und kontrovers diskutiert. Ein Teil der Wissenschaftler³ bestätigt Emotionaler Intelligenz den Status eines gut validierten und mittlerweile etablierten Konstrukts (MacCann, Joseph, Newman, & Roberts, 2014; Mayer, Caruso, & Salovey, 2000; Palmer, Gignac, Manocha, & Stough, 2005). Doch aufgrund von Unklarheiten in Konzeption, Messung und Validierung melden sich auch Kritiker zu Wort, die den Nutzen und Mehrwert oder sogar die Existenz eines entsprechenden Konstrukts anzweifeln (Locke, 2005; Schuler, 2002; van der Zee, Thijs, & Schakel, 2002). Trotz allem – oder gerade deshalb – steigt die Anzahl der Publikationen zu Emotionaler Intelligenz regelmäßig an und auch das Interesse der breiten Öffentlichkeit ist ungebrochen. „*What is this thing called emotional intelligence?*“ – mit dieser Frage beginnen nicht nur das Buch von Matthews et al. (Matthews, Emo, Roberts, & Zeidner, 2006) und das erste Kapitel der vorliegenden Arbeit, diese entscheidende Frage begleitet letztlich das gesamte Dissertationsvorhaben.

Im ersten Kapitel findet zunächst eine kurze Vorstellung des Konstrukts Emotionale Intelligenz statt. Anschließend geht es um die Messung von Emotionaler Intelligenz und damit verbundene Schwierigkeiten. Obwohl Emotionale Intelligenz in der Definition als eine Form der Intelligenz beschrieben wurde (Salovey & Mayer, 1990) – und somit als eine kognitive Fähigkeit, die idealerweise mittels Leistungstest und nicht mittels Fragebogen erfasst wird – existieren zur Erfassung derselben nach wie vor nur wenige Leistungstests. Vier Leistungstests werden in Abschnitt 1.3 vorgestellt, die Validität der Verfahren wird in Kapitel 1.4 beleuchtet. Der neue Leistungstest ViTEIP (Videobasierter Test zur Erfassung der Emotionalen Intelligenz im Pflegeberuf), der in der vorliegenden Arbeit zum Einsatz kommt, wird unter Punkt 1.5 präsentiert.

1.1 Definition und Modelle von Emotionaler Intelligenz

„*Which EI⁴ are we talking about?*“ – diese Frage formulierte Mayer auf den ersten Seiten des Buchs „Emotional Intelligence in everyday live“ (Ciarrochi, Forgas, & Mayer, 2001, p. 3). Diese Frage ist berechtigt, denn hinter dem Begriff Emotionale Intelligenz verbergen sich unterschiedliche Grund-

² Matthews, Emo, Roberts, & Zeidner (2006)

³ Zur besseren Lesbarkeit wird in der vorliegenden Arbeit ausschließlich das grammatikalische Maskulinum verwendet. Gemeint sind jedoch stets beide Geschlechter.

⁴ EI= emotional intelligence (Anmerkung der Autorin)

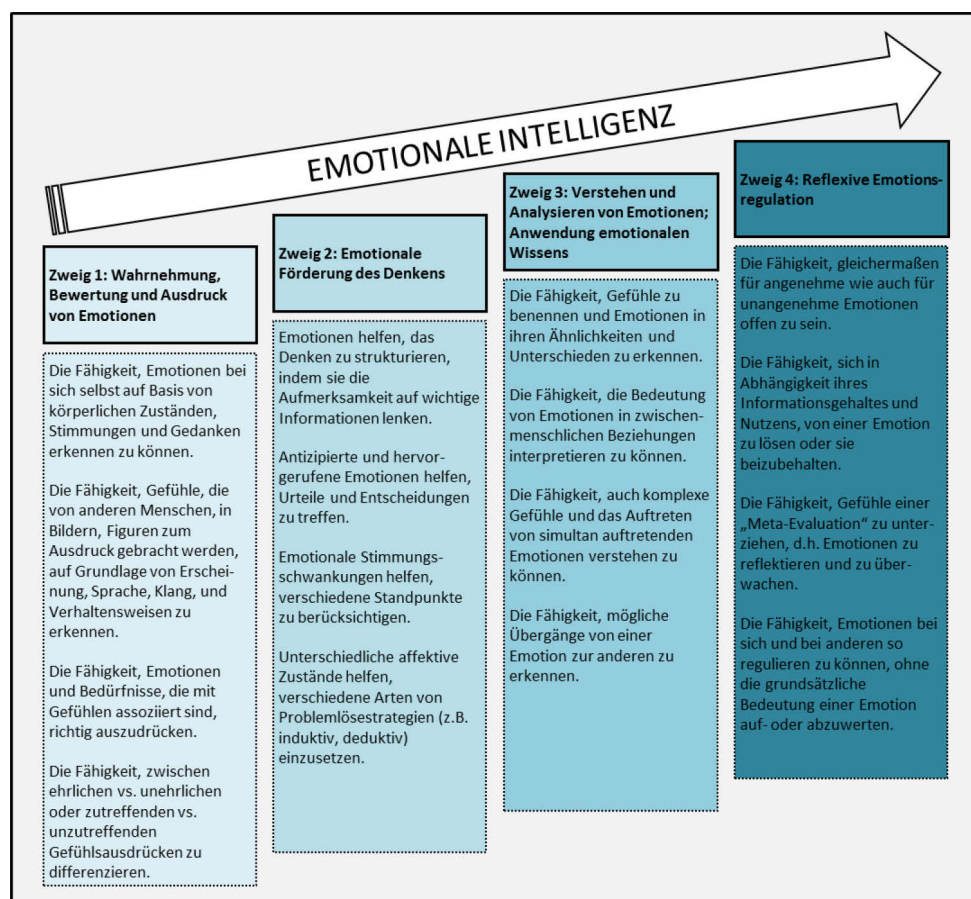
richtungen der Konzeptualisierung. Zum einen gibt es Autoren wie Goleman (1995) oder Bar-On (2000), die unter Emotionale Intelligenz ein Konglomerat verschiedener Persönlichkeitsfacetten verstehen, die wesentlich zum Erfolg in beruflichen wie privaten Lebensbereichen beitragen sollen. Diese sogenannten „Trait-Ansätze“ erfahren einige Kritik, da sie durch die breite Konzeptualisierung Emotionale Intelligenz in die Nähe von Persönlichkeitsmerkmalen, Charaktereigenschaften oder Dispositionen rücken (Ciarrochi, Chan, & Caputi, 2000; Matthews, Zeidner, & Roberts, 2012; Neubauer & Freudenthaler, 2006). Durch nicht-kognitive Aspekte wie beispielsweise Stresstoleranz, emotionale Selbstaufmerksamkeit, Bestimmtheit, Optimismus oder Impulskontrolle in Bar-On's „Modell der emotionalen und sozialen Intelligenz“ (2000) wird die Bezeichnung emotionale *Intelligenz* quasi ad absurdum geführt.

Zum anderen existieren sogenannte „Fähigkeitsmodelle“, deren Definition insgesamt weniger umstritten ist. Prominente Vertreter eines fähigkeitsbasierten Ansatzes sind Mayer und Salovey (1997), die in ihrer Konzeptualisierung die ausschließlich kognitiven Aspekte von Emotionaler Intelligenz betonen: *„Emotional Intelligence involves the ability to perceive accurately, appraise, and express emotion; the ability to access and/or generate feelings when they facilitate thought; the ability to understand emotion and emotional knowledge, and the ability to regulate emotions to promote emotional and intellectual growth“* (Mayer & Salovey, 1997, p. 10). Mayer und Salovey postulieren vier, in ihrer Komplexität aufsteigende „Zweige“ emotionaler Intelligenz, die wiederum durch jeweils vier emotionsbezogene Fähigkeiten (Facetten) definiert werden (siehe *Abbildung 1*). Die hierarchische Struktur verdeutlicht die Annahme der Autoren bezüglich der Entwicklung von Emotionaler Intelligenz, ausgehend von einfachen psychologischen Prozessen bis hin zu höher organisierten und integrativen Fähigkeiten in Zweig 4. Das Modell von Mayer und Salovey ist theoretischer Ausgangspunkt für viele Untersuchungen und Testverfahren (MacCann & Roberts, 2008; Neubauer & Freudenthaler, 2006; Sharma, Gangopadhyay, Austin, & Mandal, 2013).

In den Metaanalysen von Van Rooy, Viswesvaran und Pluta (2005) bzw. Joseph und Newman (2010) korrelierten Messungen basierend auf Fähigkeits- bzw. Trait-Modellen zu maximal $r=0.14$ und können somit als so gut wie unabhängig voneinander angesehen werden. Eine weitere Forschungsrichtung besteht aus Ansätzen⁵, die auf Informationsverarbeitungsprozesse oder Verarbeitungsstile von emotionalen Informationen fokussieren (Wilhelm, 2005). So formulieren beispielsweise Fiori und Antonakis (2012), dass es für das Verständnis von Emotionaler Intelligenz wesentlich sei, auch die zugrundeliegenden Prozesse zu untersuchen. McCrae (2000) schlägt zudem vor, sich weniger der Diskussion zu widmen, ob Emotionale Intelligenz eine kognitive Fähigkeit oder ein Trait im Sinne

⁵ In der vorliegenden Arbeit wird auf die unterschiedlichen Emotionale-Intelligenz-Modelle nicht weiter eingegangen; ein guter Überblick findet sich bei Schulze, Freund und Roberts (2006).

einer zeitlich stabilen Persönlichkeitseigenschaften sei, sondern zu untersuchen, welche spezifischen Fähigkeiten und Persönlichkeitsmerkmale bei der Verarbeitung emotionaler Informationen involviert bzw. förderlich sind. Roberts und Kollegen appellieren an die Forschungsgemeinde, sich zukünftig weniger den persönlichkeitskonzeptnahen Emotionale-Intelligenz-Ansätzen zu widmen, sondern „...im Interesse der Weiterentwicklung des Forschungsfeldes [...] Informationsverarbeitungs- und Emotionswissensmaße weiter zu entwickeln. [...] Hier bieten sich multimediasbasierte Testverfahren an, die alternative Scoringmethoden aufweisen und unterschiedliche Teilaspekte der Emotionalen Intelligenz erfassen könnten.“ (Roberts R. D., et al., 2006, S. 335). Wie in Kapitel 1.5 beschrieben wird, wurde diesem Appell bei der Entwicklung des ViTEIP (Videobasierten Test zur Erfassung der Emotionalen Intelligenz im Pflegeberuf) Rechnung getragen, indem ein videobasierter Leistungstest konzipiert wurde, der sich auf das Fähigkeitskonzept gemäß Mayer und Salovey (1997) bezieht, einen Teilaspekt von Emotionaler Intelligenz erfasst und das bislang selten eingesetzte Targetscoring verwendet werden kann.

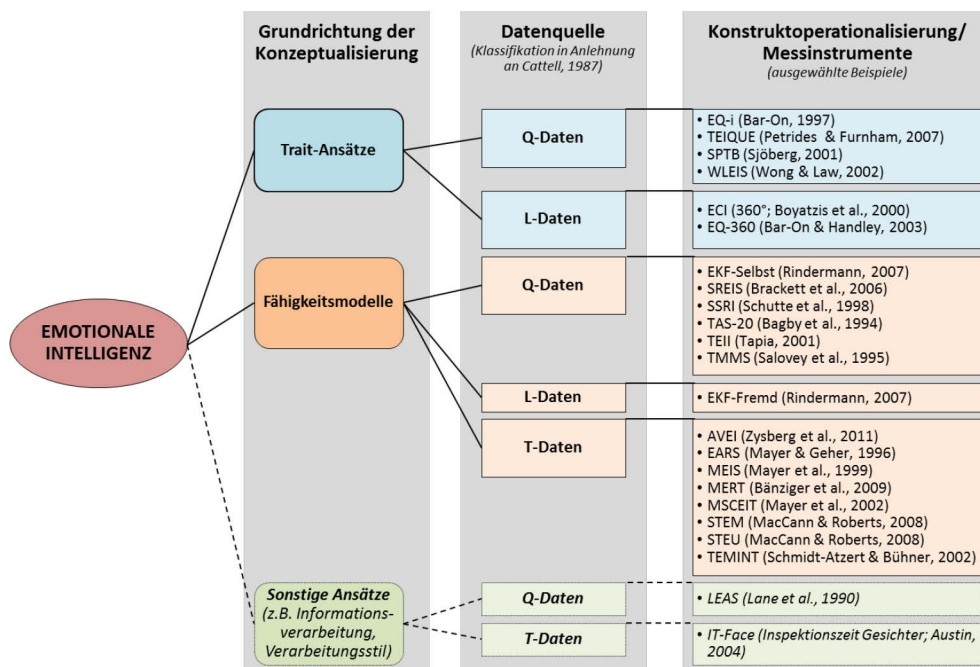


Quelle: In Anlehnung an Neubauer und Freudenthaler (2006)

Abbildung 1: Vereinfachte Darstellung des Fähigkeitsmodells der Emotionalen Intelligenz nach Mayer & Salovey (1997)

1.2 Die Messung von Emotionaler Intelligenz⁶

Neben der Klärung der theoretischen Konzeption von Emotionaler Intelligenz (d.h. Traitansatz oder Fähigkeitsmodell?) bedarf es einen genauen Blick auf die zur Operationalisierung verwendete Datenquelle. Der Klassifikation von Cattell (1978) folgend, lassen sich drei grundlegende Datenquellen identifizieren, die im Zusammenhang mit Emotionaler Intelligenz zum Einsatz kommen: sogenannte Q-Daten („questionary data“) aus Selbstbeschreibungen, T-Daten („test data“) aus Leistungstests oder Verhaltensbeobachtungen bei Arbeitsaufgaben, sowie in spezifischen Kontexten bzw. lebensgeschichtlichen Ereignissen objektiv erfassbare oder mit Fremdbeurteilungen erhobene L-Daten („life records“).



Anmerkungen:

Eine Übersicht über die Langnamen der Tests findet sich im Abkürzungsverzeichnis. Für weiterführende Informationen sei auf die Testmanuale verwiesen.

Zuordnung der Verfahren durch die Autorin; es besteht kein Anspruch auf Vollständigkeit. Aufgeführt sind auch Verfahren, die zwar nicht explizit als Tests zur Erfassung von Emotionaler Intelligenz betitelt sind, jedoch Fähigkeiten erfassen, die zu diesem Konstrukt zählen.

Abbildung 2: Schematische Darstellung der unterschiedlichen Forschungsrichtungen und Umsetzung in Messinstrumente zu Emotionaler Intelligenz

Kritisch anzumerken ist, dass sich verschiedene Autoren zwar explizit auf das Fähigkeitsmodell von Mayer und Salovey beziehen, die Operationalisierung jedoch in Form eines Fragebogens zur

⁶ Aufgrund des inhaltlichen Fokus der vorliegenden Arbeit findet eine Konzentration auf Fragen und Verfahren statt, die im Kontext von fähigkeitsbasierter Emotionaler Intelligenz zu sehen sind. Ein guter Überblick über die Messung von Emotionaler Intelligenz als Trait findet sich bei Perez, Petrides und Furnham (2006).

Selbsteinschätzung erfolgt, wie beispielsweise im „Tapia Emotional Intelligence Inventory“ (Tapia, 2001) oder „Schutte Self-Report Inventory“ (Schutte, et al., 1998). Aus der Intelligenzforschung ist hinreichend bekannt, dass Selbstbeschreibungsverfahren nur eine geringe bis maximal moderate Korrelation (um $r=0.30$) mit dem tatsächlichen Intelligenzniveau aufweisen (Furnham, 2001; Jacobs, Szer, & Roodenburg, 2012; Paulhus, Lysy, & Yik, 1998; Visser, Ashton, & Vernon, 2008); ähnlich sieht auch die Befundlage in Bezug auf Emotionale Intelligenz aus (Bowman, Markham, & Roberts, 2001; Zeidner, Shani-Zinovich, Matthews, & Roberts, 2005). Bei Emotionaler Intelligenz scheint es sogar noch schwieriger zu sein, über Fragebogen die tatsächliche Fähigkeitsausprägung zu erheben: Geht man davon aus, dass zu einem hohen Emotionale-Intelligenz-Niveau auch Reflexions- und Introspektionsfähigkeiten über die eigenen Emotionen und den Umgang damit gehören, ist anzuzweifeln, ob Personen mit einer geringen Fähigkeitsausprägung in der Lage sind, ihr eigenes Emotionale-Intelligenz-Niveau adäquat einzuschätzen (Brackett & Mayer, 2003).

Abbildung 2 soll verdeutlichen, dass verschiedenen Verfahren, die sich für die Messung von Emotionaler Intelligenz anbieten, auf unterschiedliche theoretische Grundverständnisse zurückzuführen sind und sich zudem in der verwendeten Datenquelle unterscheiden können. Eine Interpretation von Studienergebnissen sollte daher unter Berücksichtigung dieser Aspekte erfolgen.

1.2.1 Auswertungsstrategien⁷ von Emotionale-Intelligenz-Tests

Sowohl für Leistungstests im Allgemeinen als auch für Intelligenztests im Speziellen ist es wesentlich, dass Items mit klar definierten, korrekten Antworten und ein entsprechendes Auswertungs- und Bewertungsprozedere existieren (Austin & Saklofske, 2006; Guttman & Levy, 1991). Für Fragebogenverfahren, bei denen häufig der Grad der Zustimmung/Ablehnung zu einer Aussage erhoben wird, ist diese Bestimmung in der Regel unproblematisch: zu definieren ist, ob eine Zustimmung oder eine Ablehnung der Aussage für einen hohen Ausprägungsgrad des zu erfassenden Konstrukts steht.

Die Schwierigkeit bei Leistungstests zur Erfassung von alternativen Intelligenzkonzepten wie Emotionaler Intelligenz liegt darin, wie der Korrektheitsgrad einer Probandenantwort bestimmt und somit eine Leistungsbewertung vorgenommen werden kann. Im Gegensatz zu Aufgaben in klassischen Intelligenztests, bei dem objektiv richtige Antworten existieren (z.B. „ $2+3=...$?...“ oder „Wald:Bäume = Wiese:...?“⁸), ist diese Festlegung bei der Erfassung von Emotionaler Intelligenz ungleich schwieriger. In der Literatur werden für Emotionale-Intelligenz-Leistungstests drei

⁷ Die Bezeichnung Auswertungsstrategie, -prozedur, -ansatz, -methode, -prinzip und Scoring werden in der vorliegenden Arbeit synonym verwendet und beziehen sich auf die Systematik, wie der Testkennwert der Probanden berechnet wird.

⁸ Richtige Antworten: „ $2+3=5$ “ und „Wald:Bäume = Wiese:Gräser“

Methoden vorgeschlagen, wie die richtige Antwort generiert werden kann (Mayer & Geher, 1996): das Konsens-, das Experten- und das Targetscoring. Der wesentliche Unterschied der verschiedenen Ansätze liegt darin, welcher Person/Personengruppe die Kompetenz zugesprochen wird die Korrektheit einer Antwort festzulegen. Bei Konsensscoring wird über die Antwortverteilung in einer Vergleichsstichprobe die richtige Antwort definiert und bei Expertenscoring liegt diese Entscheidung bei Fachexperten für das zu erfassende Konstrukt. Das Targetscoring kann eingesetzt werden, wenn ein Test aus situativen Fragen besteht, die sich auf tatsächliche Begebenheiten beziehen. Die richtige Antwort wird von der Person, die die Situation erlebt hat, vorgegeben. Nachfolgend werden alle drei Ansätze beschrieben sowie problematische Aspekte in der Anwendung dargestellt.

Konsensscoring. Mayer und Geher (1996) führen den Ansatz des Konsensscorings im Zusammenhang mit Emotionaler Intelligenz auf. Das Rationale dieser Scoringprozedur liegt darin, dass Fähigkeits- und Wissensbereiche existieren können, in denen keine Standards für den Korrektheitsgrad einer Antwort vorliegen (Wilhelm, 2005). Das Konsensscoring ist daher nach Legree (1995) in erster Linie für Themen und Konstrukte geeignet, in denen es (noch) keine ausgewiesenen Experten oder kein klar abgrenzbares objektives Wissen gibt. Grundannahme des Konsensscorings ist, dass sich in einer großen Stichprobe mittels der Konvergenz der Probandenantworten auf eine Antwortoption die richtige Antwortalternative identifizieren lässt. Konkret bedeutet dies, dass eine Referenzstichprobe daraufhin untersucht wird, welche Antwortoption von den meisten Probanden als korrekt eingeschätzt wird. Anhand dieser Mehrheitsmeinung wird anschließend der Korrektheitsgrad der Antworten festgelegt. In der Übersicht von Legree et al. (Legree, Psotka, Tremble, & Bourne, 2005) findet sich eine ausführliche Beschreibung, unter welchen Bedingungen ein konsensbasiertes Scoring angewandt werden kann und wann von diesem Ansatz abgesehen werden sollte.

Es werden verschiedene Möglichkeiten zur Berechnung des Konsensscorings vorgeschlagen, wie beispielsweise über einfache oder standardisierte Distanzwerte, quadrierte Differenzen oder Korrelationen (jeweils zwischen Konsensmeinung und der individuellen Antwort des Probanden) (Keele & Bell, 2009; Legree, Psotka, Tremble, & Bourne, 2005). MacCann et al. (MacCann, Roberts, Matthews, & Zeidner, 2004) empfehlen, das Konsensscoring mittels Anteils- oder Modalwerten zu berechnen. Beim Vorgehen Konsens-Anteil erfolgt auf Grundlage der einzelnen Antworthäufigkeiten eine Berechnung von Anteilswerten, aus denen sich die individuellen Testkennwerte generieren (Barchard, Hensley, & Anderson, 2013; MacCann, Roberts, Matthews, & Zeidner, 2004; Mayer J. D., Salovey, Caruso, & Sitarenios, 2003). Wählten beispielsweise 45% der Probanden die Antwortoption A, 20% der Probanden die Antwortoption B und 35% die Antwortoption C, erhält

jeder Proband, der sich für die Antwort A entscheidet, einen Testkennwert von 0.45. Die Probanden, die Antwortoption B gewählt haben, erhalten einen Wert von 0.20 und diejenigen mit Antwort C 0.35. Bei der Scoringvariante Konsens-Modal wird nur diejenige Antwort als richtig gewertet und mit einem Testwert von 1 bewertet, die in der Normgruppe die höchste Zustimmung erhält (Geher, Warner, & Brown, 2001). In dem beschriebenen Beispiel träfe dies auf Antwort A zu.

Über die Jahre hinweg etablierte sich das Konsensscoring als *state-of-the-art*-Ansatz für Tests zur Erfassung von Emotionaler Intelligenz – trotz diverser Kritikpunkte. Ein maßgeblicher Kritikpunkt am Konsensscoring ist der Umgang mit schwierigen Items, was an einem Beispiel verdeutlicht werden soll. Angenommen, in einem Test mit jeweils mehreren Antwortoptionen existierte ein Item, dessen richtige Beantwortung (z.B. Wahl der Antwortoption A) nur dem obersten Zehntel der Referenzstichprobe (d.h. die Probanden mit besonders hoher Fähigkeitsausprägung) gelänge. Alle anderen Probanden entschieden sich für eine der anderen drei Antwortoptionen B (50% der Probanden), C (25%) oder D (15%). Nach der Logik des Konsensscoring ergeben sich folgende Gewichtungen: A: 0.10, B: 0.50, C: 0.25 und D: 0.15. Hier wird deutlich, dass 90% der Probanden einen höheren Wert in ihren Gesamtwert eingerechnet bekommen als die 10% der Probanden, die objektiv betrachtet die richtige Antwort gegeben und eigentlich eine höhere Leistungsfähigkeit gezeigt haben. Dies widerspricht der grundlegenden Logik von Intelligenzdiagnostik, bei der Personen mit einer hohen Intelligenz höhere Testwerte erhalten als diejenigen mit einer geringeren Leistungsfähigkeit. Die Anwendung des Konsensscorings funktioniert nach Barchard et al. (Barchard, Hensley, & Anderson, 2013) demnach nur für einfache Items, bei denen davon auszugehen ist, dass die Mehrheit der Probanden die richtige Antwort identifizieren kann.

Die Grundannahme des Konsensscoring, dass eine Übereinstimmung mit einer Referenzgruppe oder vorherrschenden kulturellen Normen als „intelligent“ angesehen wird, ist kritisch zu betrachten (Matthews, Roberts, & Zeidner, 2004), denn eine Abweichung von der Norm kann theoretisch eine herausragende oder richtige Leistung sein, die hier nicht anerkannt wird (Brody, 2004). Die höchste Punktzahl erhalten somit nicht die Probanden mit der höchsten Leistungsfähigkeit, sondern diejenigen Probanden, die konform mit der Mehrheit antworten – ein Vorgehen, das für einen klassischen Intelligenztest kaum vorstellbar wäre. Zeidner, Matthews und Roberts (2001, p. 268) sprechen in ihrem Beitrag provokativ von *Emotionaler Dummheit*, die mit konsensbasierten Auswertungen von Tests zur Erfassung von Emotionaler Intelligenz gemessen würde, anstatt vor allem am oberen Ende der Skala zu differenzieren. Je nach Stichprobenzusammensetzung können zudem durch kulturelle oder gruppenbedingte Normen, Unterschiede in der Beurteilung der Korrektheit der Antworten entstehen (Austin & Saklofske, 2006; Barchard & Russell, 2006).

Expertenscoring. Für das Expertenscoring legen Personen mit ausgewiesener Expertise in dem zu untersuchenden Gebiet fest, welche Antwort als richtig zu bewerten ist. Die Probanden erhalten für die Antwort einen Punkt, die der Antwort der Experten entspricht. Die Schwierigkeit besteht darin, dass für das Expertenscoring keine festgelegten Kriterien existieren, welche Personen als Experten in dem zu untersuchenden Bereich fungieren sollen bzw. können. Wie setzt sich eine Experten-Gruppe für Emotionale Intelligenz zusammen? Sind beispielsweise Mitglieder der „International Society of Research in Emotion“ eine adäquate Expertengruppe, so wie es für den Mayer-Salovey-Caruso Emotional Intelligence Test (Mayer, Salovey, & Caruso, 2002) angenommen wird? Reicht die theoretische Auseinandersetzung mit Emotionen aus, um ein Experte in Emotionaler Intelligenz zu sein? Einige Autoren kritisieren diese Annahme (Barchard, Hensley, & Anderson, 2013; Maul, 2012). Matthew, Zeidner und Roberts (2002) vermuten, dass selbst Experten bei der Beurteilung von anspruchsvollen, komplexen Situationen an ihre Grenzen stoßen. Gleichzeitig gibt es Studien, in denen von hohen Übereinstimmungen ($r \geq 0.88$) zwischen Konsens- und Expertenscoring berichtet wird (Follesdal & Hagtvat, 2009; Mayer, Salovey, & Caruso, 2002; Palmer, Gignac, Manocha, & Stough, 2005), woraus sich die Schlussfolgerung ableiten lässt, dass Experten fast identisch wie Laien antworten. Dies stellt die Verschiedenheit der Ansätze Konsens- und Expertenscoring, die Existenz von Experten zu Emotionaler Intelligenz und den Nutzen der Expertenauswertung in Frage.

Targetscoring. Das Target- oder Zielpersonenscoring geht im Zusammenhang von Emotionaler Intelligenz auf die frühen Arbeiten von Mayer und Geher (1996) zurück. Die Autoren beziehen sich auf die Annahme, dass nur die Person, die die Situation selbst erlebt, auch einen direkten Zugang zu ihren Emotionen hat. Die Verwendung des Targetscoring hat aus ihrer Sicht den Vorteil „...direkt und praktisch...“ zu sein (Mayer & Geher, 1996, p. 92). Für das Targetscoring wird die Person, die die Informationen (z.B. durch Schilderung einer erlebten Situation) für das Item liefert, nach ihrer Einschätzung/Empfindung befragt und daraus die richtige Antwort bestimmt. Die Auswertung erfolgt über Abweichungs- bzw. Übereinstimmungswerte zwischen der Angabe der Originalperson und der Antwort des Probanden.

Die Anwendung des Targetscorings setzt voraus, dass die Originalperson sowohl verfügbar als auch physisch, psychisch und kognitiv in der Lage ist, adäquate Angaben über die Situation und ihre eigenen Emotionen zu geben. Es besteht die Gefahr, dass die Zielperson sozial erwünschte oder erinnerungsverzerrte Antworten gibt, die nicht der tatsächlichen Begebenheit entsprechen (Mayer & Geher, 1996). Zudem fließen in die Bewertung der jeweiligen Situation und die daraus folgende Intensität der Emotion intraindividuellen Unterschiede ein. Ein- und dieselbe Situation wie beispielsweise eine Prüfung, kann selbst bei derselben Person zu unterschiedlichen Emotionen führen (z.B. Angst vor der Prüfungssituation oder Vorfreude auf eine nach der Prüfung anstehende

Feier) oder auch in der Intensität der Emotion variieren (z.B. geringe vs. große Angst). Die Antwort ist somit volatil und abhängig von Personen- und Situationsmerkmalen und der subjektiven Einschätzung der Originalperson (Siemer, Mauss, & Gross, 2007).

Bewertung und Weiterentwicklungen der Auswertungsstrategien. Welches Scoring ist nun „das beste“ für Emotionale-Intelligenz-Tests? Die Beantwortung dieser Frage ist schwierig, da – wie beschrieben – alle drei Ansätze für die Anwendung im Kontext der Erfassung von Emotionaler Intelligenz sowohl Vor- als auch Nachteile aufweisen und sich in der Literatur jeweils Befürworter und Kritiker zu Wort melden. Mayer et al. (2000) schlussfolgern aus ihrer Untersuchung mit der MEIS, dass aufgrund der höheren internen Konsistenzen, einer klareren Faktorenstruktur und höheren Korrelationen mit externen Kriterien, die Konsensauswertung den beiden anderen Ansätzen vorzuziehen sei. Auch Geher, Warner und Brown (2001) empfehlen, den Test EARS mit Konsens auszuwerten, da hier höhere Reliabilitäten erzielt würden und sich prädiktive Validität bezüglich Empathie zeige; die Auswertung mit Target sei aufgrund der möglichen Ungenauigkeit des Selbstberichts durch Antworttendenzen und mangelnder Selbstkenntnis kritisch. Da die EARS-Ergebnisse bei Auswertungen mittels Konsens und Target unkorreliert ($r=0.02$) sind, äußern Geher et al. (2001) die Vermutung, dass Emotionaler Intelligenz zwei verschiedene Fähigkeiten zugrunde liegen könnten – zum einen die Fähigkeit, bei einer Zielpersonen Emotionen einzuschätzen und zum anderen die Fähigkeit einzuschätzen, welche Emotionen andere Personen bei der Zielperson wahrnehmen. Von sehr hohen Übereinstimmungen zwischen Konsens- und Expertenscoring im MSCEIT V2.0 berichten Mayer et al. (2003) und Palmer et al. (Palmer, Gignac, Manocha, & Stough, 2005). Hier liegen die Korrelationen nach Angaben der Autoren über $r \geq 0.86$ (auf Subtest-, Skalen- und Gesamttestebene). Im Vergleich zur Konsensgruppe erwiesen sich Experten jedoch als zuverlässiger (höhere Reliabilität) und die Konvergenz auf eine Antwortoption war deutlicher (Mayer J. D., Salovey, Caruso, & Sitarenios, 2003). Mit zunehmender Verbreitung des MSCEIT setzten sich die Konsens- und Expertenauswertung als präferiertes Scoring durch, obwohl vor allem an der Verwendung der Konsensmeinung die Kritik ungebrochen ist (Brody, 2004; Roberts, Zeidner, & Matthews, 2001). Was die Autoren des MSCEIT dazu bewogen hat, sich vom ursprünglich von ihnen für die MEIS präferierten Targetscoring abzuwenden, konnte in einer persönlichen Korrespondenz mit John Mayer nicht in Erfahrung gebracht werden (Mayer, persönliche Korrespondenz im Juli 2011).

Um Unzulänglichkeiten des Konsensansatzes zu überwinden, werden Versuche unternommen alternative Auswertungsansätze zu generieren. Eine Möglichkeit liegt in der deduktiven Testkonstruktion, wo die Korrektheit einer Antwort aus der Übereinstimmung mit der Theorie abgeleitet werden kann (MacCann & Roberts, 2008). MacCann und Roberts entwickelten die Items

für den Situational Test of Emotional Understanding (STEU, siehe Kapitel 1.3.3) auf Grundlage der Theorie zur Struktur der Emotionen nach Roseman (z.B. Roseman & Evdokas, 2004) und leiteten zugleich aus dieser Theorie den Auswertungsschlüssel ab. Auch eine Kombination von Experten- und Konsensscoring kommt mittlerweile zum Einsatz, bei der die von einer Expertengruppe am häufigsten genannte Antwortalternative als richtige Antwort gewertet wird (Barchard, Hensley, & Anderson, 2013; Mayer J. D., Salovey, Caruso, & Sitarenios, 2003; Palmer, Gignac, Manocha, & Stough, 2005). In Studien der Arbeitsgruppe um Süß in Magdeburg wurde das Expertenkonsens bei der Erfassung von Sozialer Intelligenz, einem zu Emotionaler Intelligenz verwandten Konstrukt, erfolgreich eingesetzt (Baumgarten, persönliche Kommunikation am 14.12.2011). Die Frage nach der Angemessenheit der Verwendung von Mehrheitsmeinungen als Indikator für die richtige Antwort bleibt jedoch auch bei der Variante des Expertenkonsenses bestehen.

Aus Sicht der Autorin ist es erforderlich, alternative Ansätze zum Konsensscoring zu entwickeln sowie das Targetscoring weiter zu untersuchen. Denn aus der vorliegenden Befundlage ist nicht nachzuvollziehen, warum die Targetauswertung im Rahmen von Emotionaler Intelligenz so schnell an Bedeutung verloren und sich das Konsens- bzw. Expertenscoring durchgesetzt haben.

1.3 Beispiele für Leistungstests zur Erfassung von Emotionaler Intelligenz

Obwohl es mittlerweile gut 25 Jahre her ist, dass Emotionale Intelligenz in den Fokus der Aufmerksamkeit geraten ist, gibt es zur Erfassung von Emotionaler Intelligenz (verstanden als Fähigkeit) nach wie vor in erster Linie Fragebogenverfahren und nur wenige Leistungstests. Aus der Multifactor Emotional Intelligence Scale (Mayer, Caruso, & Salovey, 2000), dem ersten Leistungstest, der sich explizit auf einen fähigkeitsbasierten Ansatz von Emotionaler Intelligenz bezieht, wurde von der Arbeitsgruppe um Mayer der Nachfolgetest MSCEIT entwickelt. Da dieser Test in vielen Studien zu Emotionaler Intelligenz eingesetzt wird, soll er nachfolgend vorgestellt werden. Als Alternativen zum MSCEIT bieten sich derzeit der deutschsprachige TEMINT (Schmidt-Atzert & Bühner, 2002) sowie in englischer Sprache der STEM und STEU von MacCann und Roberts (2008) an. Auch diese Tests werden kurz vorgestellt. Ergebnisse zur Validität der Tests werden an dieser Stelle nicht aufgeführt, sondern aggregiert unter Kapitel 1.4 im Zusammenhang mit der Darstellung empirischer Befunde zu Emotionaler Intelligenz beschrieben.

Der Bedarf nach und das Interesse an Emotionale-Intelligenz-Leistungstests scheint grundsätzlich vorhanden zu sein, da nicht nur im Kontext der vorliegenden Arbeit, sondern auch in anderen Forschergruppen entsprechende Testentwicklungen vorgenommen werden. Da zum aktuellen Zeitpunkt bisher nur je eine Veröffentlichung zum „Situational Judgment Test zur Erfassung von

Emotionaler Intelligenz“ von Sharma, Gangopadhyay, Austin und Mandal (2013) sowie dem „Situational Judgment Tests of Emotional Abilities“ von Fallon et al. (Fallon, et al., 2014) existieren, wird auf eine Vorstellung dieser noch sehr jungen Verfahren verzichtet.

1.3.1 Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT)

Mit dem Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) entwickelten Mayer und Kollegen (Mayer, Salovey, & Caruso, 2002) einen Leistungstest, der mit insgesamt acht Subtests die vier Zweige „Wahrnehmung, Bewertung und Ausdruck von Emotionen“, „Emotionale Förderung des Denkens“, „Verstehen und Analysieren von Emotionen, Anwendung emotionalen Wissens“ und „Reflexive Emotionsregulation“ ihres Emotionale-Intelligenz-Modells erfassen soll (siehe *Abbildung 1*). Die Items wurden nach Angabe der Autoren aus dem Modell abgeleitet; eine kurze Beschreibung der Aufgaben ist in *Abbildung 3* aufgeführt. Die Beantwortung der Items erfolgt auf fünfstufigen Ratingskalen, deren Skalenstufen je nach Aufgabentypus unterschiedliche Bedeutungen erhalten, z.B. „gar nicht“ bis „sehr stark“ bei der Aufgabe Gesichter, „nicht ähnlich“ bis „sehr ähnlich“ bei der Aufgabe Sinneseindrücke. Die Auswertung der Items erfolgt mittels Konsens- und/oder Expertenscoring. Die Expertengruppe setzte sich zusammen aus 21 Mitgliedern der International Society of Research in Emotion; für das Konsensscoring wurde die Normierungsstichprobe mit $N=2112$ Probanden herangezogen. Aus den insgesamt 141 Items werden Kennwerte für die Bereiche erfahrungsorientierte und strategische Emotionale Intelligenz sowie für allgemeine Emotionale Intelligenz gebildet.

Wie aus der *Abbildung 3* zu entnehmen ist, liegen die internen Konsistenzen auf Skalenebene im Bereich zwischen $\alpha=0.55$ und $\alpha=0.88$; auf Ebene der Zweige und Bereiche fallen die Splithalf-Reliabilitäten mit Werten zwischen $r=0.76$ und $r=0.91$ hoch aus; siehe dazu beispielsweise auch Lopes et al. (2004) oder MacCann et al. (2011). Die von den Autoren angegebenen Reliabilitäten von über $r=0.91$ für den Gesamtwert liegen auch in Studien anderer Autoren in vergleichbarer Höhe (Maul, 2011; Palmer, Gignac, Manocha, & Stough, 2005). Die Retest-Reliabilitäten nach zwei Wochen werden mit $r_{tt}=0.86$ angegeben (Mayer J. D., Salovey, Caruso, & Sitarenios, 2003). Befunde zur Validität des Tests sind unter Kapitel 1.4 dargestellt.

Obwohl der MSCEIT ein häufig eingesetzter Emotionale-Intelligenz-Test ist, werden die Angemessenheit der Operationalisierung, die Berechnung des Gesamtestwertes, die Auswertungsprozeduren und geringe Reliabilitäten einiger Subtests kritisch gesehen (Brody, 2004; Keele & Bell, 2009; Wilhelm, 2005; Zeidner & Olnick-Shemesh, 2010). Auf der Homepage der US-amerikanischen Firma MHS, die den Vertrieb und die Auswertung des Tests durchführt, wird der MSCEIT mit „Outstanding validity and reliability“ beurteilt (<http://downloads.mhs.com/MSCEIT/MSCEIT.pdf>,

05.01.2015); eine unabhängige Einschätzung der Qualität des MSCEIT ist daher nicht einfach, da zusätzliche Informationen über das Auswertungsverfahren oder die Datengrundlage der statistischen Kennwerte nur eingeschränkt zu erhalten sind. Im Jahr 2011 wurde eine deutschsprachige Version des Tests veröffentlicht (Steinmayr, Schütz, Hertel, & Schöder-Abé, 2011).

	EMOTIONALE INTELLIGENZ							
Bereich	Erfahrungsbasierte Emotionale Intelligenz (Emotionales Erleben)				Strategische Emotionale Intelligenz (Emotionales Verarbeiten)			
Zweig	Wahrnehmung, Bewertung und Ausdruck von Emotionen		Emotionale Förderung des Denkens		Verstehen und Analysieren von Emotionen, Anwendung emotionalen Wissens		Reflexive Emotionsregulation	
Skala	Gesichter	Bilder	Sinneseindrücke	Erleichterung	Veränderungen	Komplexe Emotionen	Umgang mit eigenen Emotionen	Emotionen in Beziehungen
Beschreibung der Items	Wie ausgeprägt sind die angegebenen Emotionen in einem fotografierten Gesicht zu sehen?	Wie ausgeprägt sind die angegebenen Emotionen in einer Landschaftsfotografie oder in einem abstrakten Kunstwerk zu sehen?	Wie ähnlich sind aufgeführte Sinneseindrücke (z.B. kühl, blau, süß) einer bestimmten Emotion (z.B. Schuldgefühl)?	Wie hilfreich ist eine angegebene Emotion für die erfolgreiche Bewältigung einer vorgegebenen Situation? (z.B. wie hilfreich sind Trauer, Anspannung oder neutrale Stimmung um ein anspruchsvolles Rezept zu kochen?)	Wie kann sich eine emotionale Reaktion verändern? (z.B. eine Intensivierung von Ärger kann zu Zorn führen)	Welche Emotionen fügen sich zu anderen, komplexen Emotionen zusammen? (z.B. eine traurige Überraschung führt zu Enttäuschung, Verwunderung, Zorn, Furcht oder Bedauern?)	Wie wirksam sind unterschiedliche Verhaltensalternativen zur eigenen Emotionsregulation?	Wie wirksam sind unterschiedliche Verhaltensalternativen, um ein bestimmtes Ziel zu erreichen?
Anzahl der Items	4 Fotografien mit je 5 Emotionen = insgesamt 20 Items	6 Kunstwerke bzw. Fotografien mit je 5 Emotionen = insgesamt 30 Items	5 Situationsbeschreibungen mit je 3 Sinneseindrücken = insgesamt 15 Items	5 Szenarien mit je 3 Emotionen = insgesamt 15 Items	20 Items	12 Items	5 Szenarien mit je 4 Reaktionsmöglichkeiten = insgesamt 20 Items	3 Szenarien mit je 3 Reaktionsmöglichkeiten = insgesamt 9 Items
Reliabilität	Konsens: $\alpha=0.80$ Experten: $\alpha=0.82$	Konsens: $\alpha=0.88$ Experten: $\alpha=0.87$	Konsens: $\alpha=0.65$ Experten: $\alpha=0.55$	Konsens: $\alpha=0.64$ Experten: $\alpha=0.63$	Konsens: $\alpha=0.70$ Experten: $\alpha=0.68$	Konsens: $\alpha=0.66$ Experten: $\alpha=0.62$	Konsens: $\alpha=0.69$ Experten: $\alpha=0.64$	Konsens: $\alpha=0.67$ Experten: $\alpha=0.64$
	Konsens: $r_s=0.91$ Experten: $r_s=0.90$		Konsens: $r_s=0.79$ Experten: $r_s=0.76$		Konsens: $r_s=0.80$ Experten: $r_s=0.77$		Konsens: $r_s=0.83$ Experten: $r_s=0.81$	
	Konsens: $r_s=0.90$ Experten: $r_s=0.90$				Konsens: $r_s=0.88$ Experten: $r_s=0.86$			
	Konsens: $r_s=0.93$ Experten: $r_s=0.91$							

Anmerkungen:

Konsens= Auswertung mit dem Konsensscoring; Experten= Auswertung mit Expertenscoring. α = Cronbachs Alpha; r_s = Splithalf Reliabilität.

Abbildung 3: Übersicht über den MSCEIT V2.0 inkl. Angaben zur Reliabilität nach Mayer et al. (Mayer J. D., Salovey, Caruso, & Sitarenios, 2003)

1.3.2 Test zur Erfassung der Emotionalen Intelligenz (TEMINT)

Anders als im MSCEIT werden im TEMINT (Schmidt-Atzert & Bühner, 2002) nicht Emotionen in verschiedenen Materialien (Gesichter, Kunstwerke, Musik,...) dargeboten, sondern es geht um das Erkennen von Emotionen nach der Beschreibung einer spezifischen Situation. Diese Situationsvorgabe kann nach der Begriffsprägung von Schmidt-Atzert (2003) als „situativer Auslöser“ bezeichnet werden. Messgegenstand im TEMINT ist ein Teilbereich von Emotionaler Intelligenz, nämlich die Fähigkeit einer Person, in einer beschriebenen Situation anhand situativer Auslöser Emotionen einer Zielperson zu erkennen und richtig zu benennen. Die Items wurden aus realen Situationsbeschreibungen generiert, indem Personen gebeten wurden, nach dem Erleben von Alltagssituationen diese kurz zu beschreiben und Angaben über den eigenen Gefühlszustand zu machen.

Den Probanden werden insgesamt zwölf Situationsbeschreibungen schriftlich präsentiert; *Abbildung 4* zeigt eine Beispielsituation. Aufgabe ist es, auf einer dreistufigen Skala einzuschätzen, wie stark die Emotionen *Abneigung, Ärger, Angst, Unruhe, Traurigkeit, Schuldgefühle, Freude, Stolz, Zuneigung und Überraschung* bei der Zielperson in der beschriebenen Situation ausgeprägt waren. Wird jede Emotion innerhalb einer Situation als ein Item verstanden, besteht der TEMINT aus insgesamt 85 Items. Da die im TEMINT präsentierten Situationen auf wahren Begebenheiten basieren, kann die Auswertung mittels Targetscoring vorgenommen werden: die von einem Probanden angekreuzte Intensität der Emotion wird mit der Angabe der Originalperson verglichen und ein Abweichungswert berechnet. Geringere Abweichungswerte bedeuten daher im TEMINT eine bessere Testleistung als höhere Werte. Im Beispiel in *Abbildung 4* erhält der Proband für die Emotionen Abneigung, Überraschung und Schuldgefühl je einen Abweichungswert von 1, für die Emotionen Ärger, Angst, Unruhe und Trauer einen Abweichungswert von 0. Die Reliabilität des TEMINT liegt bei $\alpha=0.77$ (Schmidt-Atzert & Bühner, 2002) bzw. $\alpha=0.76$ (Amelang & Steinmayr, 2006). Validitätsbelege werden unter 1.4 aufgeführt.

Studentin, 24 Jahre: „Ich habe eine wichtige Prüfung nicht bestanden und musste sie noch einmal wiederholen.“

Versetzen Sie sich in die 24-jährige Studentin hinein.
Wie stark waren ihre Gefühle? Machen Sie bei jedem Gefühl ein Kreuz!

	nicht vorhanden oder sehr schwach	schwach bis mittel	stark bis sehr stark
Abneigung ...	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Ärger	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Angst	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Unruhe	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Traurigkeit ...	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Schuldgefühl .	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Freude	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Stolz	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Zuneigung ...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Überraschung .	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>

Anmerkungen:

☒ = Urteil des Probanden; ☒ = Angabe der Originalperson.

Abbildung 4: Beispielsituation des TEMINT inkl. Auswertung Target-Abweichung

1.3.3 Situational Test of Emotion Management (STEM) und Situational Test of Emotional Understanding (STEU)

Mit dem Situational Test of Emotion Management (STEM) und dem Situational Test of Emotional Understanding (STEU) entwickelten MacCann und Roberts (2008) zwei Leistungstests, die je eine Facette (Emotionen regulieren bzw. Emotionen verstehen) des Mayer-Salovey-Modells erfassen.

Situational Test of Emotion Management (STEM). Im Vergleich zum MSCEIT, bei dessen Entwicklung die Itemauswahl und die Entscheidung über die Richtigkeit einer Antwort zwar theoriegeleitet, aber in erster Linie durch das Urteil der Autoren geprägt ist, wählten MacCann und Roberts (2008) beim STEM eine andere, für Situational Judgment Tests typische Herangehensweise: Die Entwicklung der Items erfolgte mittels teilstrukturierter Interviews, in denen Probanden emotionale Erlebnisse der letzten 14 Tage berichteten. Anschließend wurden die Situationen nach inhaltsanalytischen Kriterien in Items transformiert. Die Antwortoptionen wurden empirisch mit einer weiteren Stichprobe sowie mit zwei Gruppen von Experten entwickelt. Dieses Vorgehen ist aus Sicht der Testautoren zwar zeit- und ressourcenaufwändig, verspricht aber eine hohe ökologische Validität, akzeptable Reliabilitäten und ermöglicht einen Einsatz in Bereichen, in denen noch unklar ist, welches die idealtypische Verhaltensweise in der jeweiligen Situation darstellt.

Jacob is having a large family gathering to celebrate him moving into his new home. He wants the day to go smoothly and is a little nervous about it. *What action would be the most effective for Jacob?*

- (a) Talk to friends or relatives to ease his worries.
- (b) Try to calm down, perhaps go for a short walk or meditate.
- (c) Prepare ahead of time so he has everything he needs available.
- (d) Accept that things aren't going to be perfect but the family will understand.

Abbildung 5: Beispielsituation des STEM (siehe: Ferguson & Austin, 2011)

Der STEM besteht aus insgesamt 44 Items, wovon sich 12 Items auf die Emotion Angst, 18 Items auf die Emotion Ärger und 14 auf die Emotion Trauer beziehen. Je ca. die Hälfte der Items enthalten Situationsbeschreibungen eines beruflichen bzw. privaten Kontexts. Aufgabe der Probanden ist es, aus vier Antwortoptionen die Alternative zu wählen, die für die beschriebene Situation am effektivsten zu sein scheint. In *Abbildung 5* ist ein Beispielitem aufgeführt. Der STEM liegt sowohl in einer Multiple-Choice-Variante als auch mit dem Antwortformat einer Ratingskala vor. Die Auswertung des STEM erfolgt über ein Expertenscoring. Die Expertengruppe ($N=19$) setzt sich zusammen aus Mitglieder eines Forschungskonsortiums zu Emotionale Intelligenz, qualifizierten Psychologen und Coaches mit Erfahrungen in Beratung oder Therapie. Als interne Konsistenzen

geben MacCann und Roberts (2008) für das Multiple-Choice-Format $\alpha=0.68$ und das Ratingskalenformat $\alpha=0.92$ an, bei Austin (2010) ergibt sich ein Wert von $\alpha=0.67$ (Multiple-Choice). Erste Ergebnisse zur Retest-Reliabilität werden von Libbrecht et al. (Libbrecht, Lievens, Carette, & Côté, 2014) mit $r_{tt}=0.85$ berichtet.

Situational Test of Emotional Understanding (STEU). Aufgrund der rational-deduktiven Konstruktionsstrategie können für den STEU sowohl die Items als auch das Scoring aus der Theorie zur Struktur der Emotionen nach Roseman (Roseman & Evdokas, 2004) abgeleitet werden. Die Theorie beschreibt, wie durch kognitive Bewertungsprozesse spezifische Situationstypen zu bestimmten Emotionen führen. Einflussgrößen dieses Modells sind die Situation (konsistent vs. inkonsistent zu den eigenen Motiven), der motivationale Zustand (appetitiv vs. aversiv), Verantwortlichkeit (Umstände der Situation vs. Andere vs. Selbst), Überraschung (erwartet vs. unerwartet), Sicherheit des Eintretens der Situation (sicher vs. unsicher), Kontrollmöglichkeit (hoch vs. gering) und Problemtypus (instrumental vs. intrinsisch). So führt beispielsweise eine „unerwünschte Situation, deren Auftreten unwahrscheinlicher wird oder sogar nicht eintritt“ nach der Theorie zu der Emotion „Erleichterung“ (nach MacCann & Roberts, 2008). Für die drei Kontextbereiche neutral/privat/beruflich wurden insgesamt 42 Situationsschilderungen generiert (Beispielitems in *Abbildung 6*); die Probanden geben ihre Antworten im Multiple-Choice-Prinzip.

Die Auswertung des Tests erfolgt über einen Vergleich zwischen der Probandenantwort und der Antwort, die sich aus der Theorie ableiten lässt. Mit der Existenz dieses eindeutigen Kriteriums setzen sich MacCann und Roberts von anderen Testautoren ab, die auf das häufig kritisierte Konsens- oder Expertenscorings zurückgreifen. Die Reliabilität wird von den Autoren mit $\alpha=0.71$ angegeben. Bei Austin (2010) und Libbrecht et al. (Libbrecht, Lievens, Carette, & Côté, 2014) hingegen fällt die Reliabilität des STEU mit $\alpha=0.48$ bzw. $\alpha=0.40$ deutlich geringer aus. Die Kurzform STEU-B kann als Screening für Emotionale Intelligenz eingesetzt werden (Allen, Weissmann, Hellwig, MacCann, & Roberts, 2014).

<p>An unwanted situation becomes less likely or stops altogether. The person involved is most likely to feel: (a) regret, (b) hope, (c) joy, (d) sadness, (e) relief</p> <p>An irritating neighbor of Eve's moves to another state. Eve is most likely to feel? (a) regret, (b) hope, (c) joy, (d) sadness, (e) relief</p> <p>A supervisor who is unpleasant to work for leaves Alfonso's work. Alfonso is most likely to feel? (a) regret, (b) hope, (c) joy, (d) sadness, (e) relief</p>
--

Abbildung 6: Beispielsituationen des STEU für verschiedene Kontexte neutral/privat/beruflich (MacCann & Roberts, 2008)

1.4 Die Validität von Emotionale-Intelligenz-Tests

Welches Konstrukt wird mit den zuvor beschriebenen Tests erfasst? Die Beantwortung dieser Frage erfolgt über Untersuchungen der faktoriellen Struktur sowie der konvergenten, diskriminanten, externen und inkrementellen Validität. Um die konvergente Validität eines Emotionale-Intelligenz-Tests zu belegen, werden enge Zusammenhänge zwischen Verfahren sowohl zum selben als auch zu konstruktnahen Konzepten (z.B. Empathie) erwartet. Demgegenüber steht die diskriminante Validität, die verbunden ist mit der Erwartung über eine höchstens moderate Korrelation zu konstruktfernen Verfahren wie beispielsweise Persönlichkeitstests. Je nach Facette werden jedoch Unterschiede in der Höhe angenommen; so ist zu der BIG-5-Komponente Neurotizismus eine geringere (sowie negative) Korrelation zu erwarten als beispielsweise zu Offenheit für Neues oder Verträglichkeit.

Wie eng darf oder soll der Zusammenhang zu psychometrischer Intelligenz ausfallen? Mayer, Caruso und Salovey (2000) erwarten von Emotionale-Intelligenz-Tests, dass sich einerseits positive Korrelationen zu etablierten Intelligenzdomänen (z.B. verbaler Intelligenz) ergeben, um die Verwandtschaft zwischen Intelligenz und Emotionaler Intelligenz zu belegen, andererseits die Korrelationen nicht so hoch sind, dass es sich um identische Konstrukte handelt. Damit kann der Zusammenhang zwischen diesen beiden Intelligenzformen sowohl unter Aspekten der konvergenten als auch diskriminanten Validität behandelt werden. Um den kognitiven Aspekt von Emotionaler Intelligenz zu betonen, wird in der vorliegenden Arbeit dieser Zusammenhang unter konvergenter Validität gesehen (Mayer, Caruso, & Salovey, 2000; Schmidt-Atzert & Bühner, 2002; Zeidner, Shani-Zinovich, Matthews, & Roberts, 2005).

Nachfolgend werden ausschließlich Ergebnisse aus Studien dargestellt, in denen Emotionale Intelligenz als Fähigkeit verstanden wird und überwiegend die oben vorgestellten Leistungstests zur Erfassung des Konstrukts eingesetzt wurden.

Faktorielle Struktur. Die Erwartung ist, dass sich die theoretische Struktur eines Konstrukts (hier: Emotionale Intelligenz) auch in Ergebnissen zugehöriger Tests wiederfindet. Mayer et al. (Mayer, Salovey, & Caruso, 2002) gehen davon aus, dass sich in Ergebnissen des MSCEIT die postulierte Vier-Faktoren-Struktur ihres Emotionale-Intelligenz-Modells (siehe Kapitel 1.1) abbildet. Einige Autoren (Day & Carroll, 2004; Livingstone & Day, 2005; Mayer J. D., Salovey, Caruso, & Sitarenios, 2003) bestätigten in ihren Studien diese Struktur. Palmer, Gignac, Manocha und Stough (2005), Rode et al. (2008) sowie Gardner und Qualter (2011) hingegen sprechen sich nach konfirmatorischer Analyse verschiedener Modelle für ein Drei-Faktoren-Modell aus, in welchem die beiden ersten Zweige „Wahrnehmung von Emotionen“ und „emotionale Förderung des Denkens“ eine gemeinsame

Komponente bilden. Ein ähnliches Ergebnis zeigte sich auch in der metaanalytischen Untersuchung von Fan et al. (Fan, Jackson, Yang, Tang, & Zhang, 2010). Unter Verwendung von Daten aus 19 Studien mit dem MSCEIT favorisieren die Autoren nicht die von Mayer und Kollegen postulierte vierfaktorielle Lösung, sondern aufgrund einer hohen Korrelation zwischen den oben genannten Zweigen eine dreifaktorielle Lösung; die angenommene Existenz eines Generalfaktors der Emotionalen Intelligenz sehen sie bestätigt. Doch es gibt auch Kritiker, die die Angemessenheit des Emotionale-Intelligenz-Modells von Mayer et al. grundsätzlich anzweifeln und damit die Konstruktvalidität des MSCEIT in Frage stellen (Spector & Johnson, 2009). Follesdal und Hagvet (2009) untersuchten an einer norwegischen Stichprobe den MSCEIT und identifizieren verschiedene Messfehlerquellen. Der Großteil der Varianz des MSCEIT werde nicht durch die Facette „Person“ erklärt, sondern durch Inkonsistenzen in den Items. Eine Generalisierbarkeit auf die angenommenen Validitätsbereiche sei zudem aufgrund der Multidimensionalität einzelner Skalen und der Schwierigkeit, diese zu interpretieren, kaum möglich. Diese Einschätzung steht der Annahme von Mayer und Kollegen entgegen, die hinter den „Zweigen“ ihres Emotionale-Intelligenz-Modells bzw. den Aufgaben des MSCEIT ein gemeinsames Konstrukt sehen.

Auch andere Tests zur Erfassung von Emotionaler Intelligenz haben mit der Konstruktvalidität zu kämpfen: Bei der Untersuchung des STEM bzw. STEU mittels explorativer Faktorenanalyse zeigten sich nicht die erwarteten faktoriellen Zuordnungen der Items (Ferguson & Austin, 2011). Die angenommene dreifaktorielle Struktur mit den Emotionen Angst, Ärger und Trauer im STEM sowie die drei Kontextbereiche neutral/privat/beruflich im STEU bestätigten sich nicht.

Konvergente Validität. Bezüglich der konvergenten Validität sind die Korrelationen zwischen unterschiedlichen Verfahren zum selben sowie konstrukt-nahen Konzepten und zu Intelligenz von Interesse. Erwartet werden enge Zusammenhänge, um die Verwandtschaft der Konstrukte zu belegen. In *Tabelle 1* sind Korrelationen aus nachfolgend zitierten Studien übersichtlich dargestellt.

In Untersuchungen bestätigen sich die erwarteten *Zusammenhängen zwischen verschiedenen Emotionale-Intelligenz-Leistungstests*. Maul (2011) setzte die beiden Tests MEIS und MSCEIT ein und fand auf Gesamttestebene hohe ($r=0.80^{**}$) Korrelationen bzw. auf Skalenebene Korrelationen bis $r=0.69^{**}$ zwischen den Verfahren. Da der MSCEIT jedoch den Nachfolgetest der MEIS darstellt, ist ein solcher Zusammenhang wenig überraschend und auch im Sinne einer erfolgreichen Überprüfung der Konzeptualisierung zu interpretieren. Die beiden Tests STEM und STEU korrelieren positiv, jedoch in unterschiedlicher Höhe. MacCann und Roberts (2008) berichten von $r=0.70^{**}$; bei Libbrecht und Lievens (2012) oder Austin (2010) fallen die Korrelationen mit $r=0.24^{**}$ bzw. $r=0.29^{**}$ deutlich geringer aus. Letztgenannte Korrelationshöhen scheint eher dabei die Regel als die

Ausnahme für Emotionale-Intelligenz-Tests zu sein: So korrelierte der MSCEIT mit dem STEU zu $r=0.33^{**}$ und dem STEM zu $r=0.36^{**}$ (Austin, 2010). Auch bei Blickle et al. (Blickle G. , Momm, Liu, Witzki, & Steinmayr, 2011) lagen die Korrelationen zwischen TEMINT und MSCEIT im vergleichbaren Bereich von $r=-0.22^{**}$ (Anmerkung: aufgrund der Berechnung von Abweichungswerten bedeuten beim TEMINT hohe Werte eine schlechte Leistung und geringe Werte eine gute Leistung).

Die aufgeführten Ergebnisse verweisen darauf, dass es Schnittmengen zwischen den von den Tests erfassten Konstrukten gibt. Dennoch ist von einer vorschnellen Einschätzung abzuraten, da auch Untersuchungen existieren, in denen sich nur minimale bzw. keine Zusammenhänge zeigen, wie beispielsweise bei Ciarrochi et al. (2003): mit dem MEIS-Gesamtscore interkorrelierte die LEAS (Level of Emotional Awareness Scale) nicht signifikant ($r=0.15$) und auf MEIS-Skalenebene zu maximal $r=0.21^*$. Keine signifikante Korrelation zwischen MSCEIT und TEMINT fand sich bei Choi und Kluemper (2012).

Bei der Beurteilung des *Zusammenhangs von Emotionaler Intelligenz mit Empathie* ist zu berücksichtigen, dass Empathie in der Regel mit Fragebögen erhoben wird, die teilweise unklare psychometrische Qualität besitzen. Zwischen der MEIS und einem Empathietest zeigten sich bei Mayer et al. (Mayer, Caruso, & Salovey, 2000) sowohl auf Gesamttest- als auf Skalenebene positive Korrelationen um $r=0.33^{**}$. Der TEMINT korrelierte mit den Ergebnissen in einem Empathie-Fragebogen zu $r=-0.26^{**}$ (Blickle G. , et al., 2009; Anm.: Durch die schon genannten Abweichungswerte des TEMINT weist die genannte Korrelation mit Empathie in die erwartete Richtung), bei Siebert (2007) ergibt sich hingegen keine signifikante Korrelation zwischen TEMINT und Empathie ($r=-0.10$ n.s.).

Die Zusammenhänge zwischen *Emotionale-Intelligenz-Tests und Verfahren zur Emotionswahrnehmung oder -verarbeitung* fallen eher heterogen aus. Für den STEU –jedoch nicht für den STEM – ergaben sich erwartungskonform signifikant positive Zusammenhänge zu verwandten Aufgaben wie z.B. Erkennen von Emotionen auf Gesichtern oder Reaktionszeit zur Identifikation von Trauer (Austin, 2010). Der TEMINT korrelierte mit Ergebnissen eines Tests zur Emotionswahrnehmung (DANVA) zu $r=0.26^{**}$ (Blickle G. , et al., 2009). Der DANVA-Gesamtwert wird wie der TEMINT invers berechnet; die Korrelation entspricht somit der erwarteten Richtung. Bei MacCann und Roberts (2008) hingegen fanden sich für STEM und STEU nur zum Teil die erwarteten Korrelationen mit den Skalen des Alexithymie-Tests TAS-20. Alexithymie bildet als Gefühlsblindheit quasi den unteren Pol des Emotionale-Intelligenz-Kontinuums ab (Parker, 2006). Roberts et al. (2006) untersuchten die Beziehung zwischen MSCEIT und verschiedenen Aufgaben zur Emotions-

wahrnehmung. Entgegen der Annahme der Emotionale-Intelligenz-Theorie, basierten die Tests zur Emotionswahrnehmung und die MSCEIT-Faktoren nicht auf einem gemeinsamen Faktor.

Die Spannweite der Korrelation von *Emotionaler Intelligenz und psychometrischer Intelligenz(-facetten)* ist groß: Sie reicht von nicht signifikanten orthogonalen Zusammenhängen bis zu signifikant positiven Korrelationen; *siehe Tabelle 1*. Für fehlende oder nur geringe Zusammenhänge seien exemplarisch die Studien von Blickle et al. (TEMINT und Wonderlic Personnel Test; Blickle, Kramer, & Mierke, 2010) oder Austin (STEM und verbale Intelligenz; Austin, 2010) genannt. Signifikante Zusammenhänge fanden sich hingegen in den Studien von Schulte et al. (MSCEIT und Wonderlic Personnel Test; Schulte, Ree, & Carretta, 2004) oder MacCann und Roberts (STEU bzw. STEU und verbale Intelligenz; 2008). Die Heterogenität der Ergebnisse zeigte sich auch in zwei Stichproben bei Farrelly und Austin (2011) mit unterschiedlichen Ergebnissen; die Autoren ziehen dennoch die Schlussfolgerung, dass der MSCEIT mit kristalliner Intelligenz in Verbindung stehe, der MSCEIT jedoch kein reines Verfahren zur Messung einer einzigen Fähigkeit sei. Brody (2009) vermutet, dass Emotionale Intelligenz eine Komponente von „g“ sei und nicht, wie zum Teil von Befürwortern des Konstruktes postuliert, eine Alternative zu „g“ darstelle. MacCann et al. (MacCann, Joseph, Newman, & Roberts, 2014) lieferten mit ihrer Untersuchung einen lange erwarteten (ersten) Beleg, dass Emotionale Intelligenz als ein Faktor des Zweiten Stratum⁹ in ein etabliertes Intelligenzmodell (Cattell-Horn-Carroll-Modell; Horn & Blankson, 2012) integriert werden kann. Aufgaben, in denen es um die Verarbeitung von emotionalen Informationen geht, bildeten einen separaten Gruppenfaktor auf der Ebene von fluider Intelligenz oder visueller Verarbeitung und ergänzten somit das bestehende Modell.

Diskriminante Validität. Wie in *Tabelle 1* dargestellt, fallen Korrelationen zwischen Verfahren zur Erfassung von *Emotionaler Intelligenz und Persönlichkeitsmerkmalen* (BIG-5) überwiegend erwartungskonform aus. Bei Schulte, Ree und Carretta (2004) lagen die Korrelationen des MSCEIT zwischen $r=-0.28^*$ (Neurotizismus) und $r=0.27^*$ (Offenheit für Neues und Verträglichkeit); bei Austin et al. (Austin, Farrelly, Black, & Moore, 2007) zwischen $r=-0.07$ (Neurotizismus) und $r=0.21^{**}$ (Verträglichkeit). Auch für STEM, STEU und TEMINT zeigte sich diskriminante Validität zu Persönlichkeit (Blickle G. , et al., 2009; MacCann & Roberts, 2008). Wie eingangs beschrieben, sind gewisse signifikante Korrelationen zwischen den Verfahren zur Erfassung von Emotionaler Intelligenz und Persönlichkeit durchaus zu erwarten, wie beispielsweise eine moderaten Korrelation

⁹ Im Cattell-Horn-Carroll-Modell wird angenommen, dass sich Intelligenz in drei hierarchisch angeordneten „Schichten“ (Strata) abbilden lässt (Flanagan & Dixon, 2014). An der Spitze (Stratum III) steht ein Generalfaktor „g“, darunter liegen auf dem breiter angelegten, zweiten Stratum spezifische Fähigkeitskomplexe (bspw. fluide Intelligenz „Gf“, kristalline Intelligenz „Gc“, Informationsverarbeitungsgeschwindigkeit „Gs“). Das Stratum I besteht aus engen Fähigkeiten wie Lesegeschwindigkeit, mathematisches Wissen, Arbeitsgedächtnis usw.

zu Verträglichkeit (Austin, Farrelly, Black, & Moore, 2007; Libbrecht & Lievens, 2012; MacCann & Roberts, 2008) oder Offenheit (Blickle G. , et al., 2009; Libbrecht & Lievens, 2012). In den Definitionen von Verträglichkeit (u.a. als das altruistische Bemühen anderen zu helfen und kooperatives Verhalten zu zeigen sowie einem ausgeprägten Harmoniebedürfnis) und Offenheit (u.a. als die Offenheit für das Erleben positiver und negativer Gefühle, Interesse an persönlichen und öffentlichen Vorgängen) zeigt sich die inhaltliche Nähe zu Emotionaler Intelligenz. Hier spiegelt sich eine Disposition wider, die sich – wie auch bei Emotionaler Intelligenz – auf den Umgang mit Emotionen bzw. anderen Menschen bezieht. Fiori und Antonakis (2011, p. 333) weisen darüber hinaus auf die unterschiedlichen Korrelationsmuster zwischen einzelnen Emotionale-Intelligenz-Skalen und verschiedenen Persönlichkeitsfacetten hin: Die Verwendung des MSCEIT-Gesamtscores verschleierte die wahren Zusammenhänge zwischen Emotionaler Intelligenz und Persönlichkeitsmerkmalen.

Externe Validität. Vor allem in populärwissenschaftlichen Medien wurde Emotionale Intelligenz als Konstrukt angepriesen, das in unterschiedlichen Lebensbereichen über Erfolg oder Misserfolg entscheide und eine bessere Vorhersage ermögliche als der IQ (Ciarrochi, Forgas, & Mayer, 2001; Goleman, 1995; TIME, 1995). Doch das Niveau, an dem sich Emotionale Intelligenz in Bezug auf Vorhersagekraft von praktisch-relevanten Kriterien messen lassen muss, ist hoch: Nach wie vor gelten kognitive Fähigkeitstests (v.a. IQ-Tests) als beste Prädiktoren für verschiedene Erfolgskriterien in Schule, Beruf und anderen Lebenssituationen (Schmidt & Hunter, 1998). In Barchards (2003) umfangreicher Untersuchung konnten verschiedene Emotionale-Intelligenz-Facetten nur annähernd 8% der Varianz an akademischen Leistungen erklären – allgemeine Indikatoren kognitiver Fähigkeiten leisten hingegen ca. 17% Varianzaufklärung.

MacCann und Roberts (2008) führen Korrelationen von $r=0.42^{**}$ bzw. $r=0.34^{**}$ zwischen der Psychologienote im Grundstudium und dem STEU bzw. STEM auf. Die Unterschiede in Korrelationsmustern auf Ebene der MSCEIT-Zweige, wie sie Fiori und Antonakis (2011, p. 333) in Bezug auf die Korrelation zu Persönlichkeitsmerkmalen beschreiben, findet sich auch zu Leistungskriterien wieder: Emotionsverständnis korrelierte zum verbalen SAT (Schulleistungstest) zu $r=0.36^{**}$, Emotionsnutzung hingegen zu $r=-0.22^{**}$. Bei der Längsschnittstudie von Blickle und Kollegen (Blickle G. , Momm, Liu, Witzki, & Steinmayr, 2011) zeigte sich, dass Emotionale Intelligenz als Moderator für den Zusammenhang von Aufstiegsmotiv und Berufserfolg fungiert: Bei Nachwuchskräften führte ein hohes Aufstiegsmotiv nur bei gleichzeitig hoher Emotionaler Intelligenz zu beruflichem Erfolg. Ein Moderatoreffekt des TEMINT zwischen psychischer Gesundheit und Berufserfolg verschwand bei Auspartialisierung von Geschlecht, Alter, wöchentlicher Arbeitszeit und Firmenzugehörigkeit (Lindebaum, 2013). Joseph und Newman (2010) vermuten, dass in Abhängigkeit von den

Anforderungen an den Umgang mit Emotionen nur in bestimmten Berufen und/oder Positionen Emotionale Intelligenz eine Rolle spielt, in anderen wiederum nicht. Inwiefern Emotionale Intelligenz beispielsweise im Zusammenhang mit Führung eine Rolle spielt, wird von verschiedenen Autoren sehr unterschiedlich beurteilt und kontrovers diskutiert (Antonakis, Ashkanasy, & Dasborough, 2009; Furnham, McClelland, & Mansi, 2012).

Jenseits der Korrelationen zu Leistungsindikatoren zeigten sich erwartungskonform einige positive Ergebnisse in Bezug auf sozial relevante Faktoren. Ciarrochi, Chan und Caputi (2000) berichteten von einem positiven Zusammenhang von Emotionaler Intelligenz (MEIS) mit Lebenszufriedenheit und selbstberichteter Zufriedenheit von Beziehungen. Emotionale Intelligenz, gemessen mit dem MSCEIT, stand in einem negativen Zusammenhang mit Psychopathologie, antisozialem Verhalten oder Drogen- und Alkoholkonsum (Brackett, Mayer, & Warner, 2004; Lishner, Swim, Hong, & Vitacco, 2011; Visser, Bay, Cook, & Myburgh, 2010) und in einem positiven Zusammenhang zu Stressverarbeitung (Schneider, Lyons, & Khazon, 2013). In den Metaanalysen von Martins, Ramalho und Morin (2010) und Schutte et al. (Schutte, Malouff, Thorsteinsson, Bhullar, & Rooke, 2007) zeigten sich moderate positive Korrelationen zwischen „Gesundheit“ und „Emotionaler Intelligenz“.

Inkrementelle Validität. Ein wesentlicher Aspekt im Kontext der Erforschung neuer Konstrukte und der Entwicklung von Tests ist die inkrementelle Validität über etablierte Konzepte (v.a. Intelligenz und Persönlichkeit) hinaus in Bezug auf Leistungsmaße oder andere Kriterien. Emotionale-Intelligenz-Tests zeigten in verschiedenen Studien das Potenzial, signifikante Anteile an Varianz bei bedeutsamen und sozial relevanten Ergebnissen über etablierte Konstrukte hinaus zu erklären, wie beispielsweise bei Lopes et al. (2003). Hier ergab die MSCEIT-Skala „Emotionen Managen“ inkrementelle Validität nach Kontrolle von Verbaler Intelligenz und den BIG-5 in Bezug auf Beziehungsqualität. Davis und Humphrey (2012) berichteten von einer Prädiktionskraft von $R^2=0.20$ des MSCEIT über Persönlichkeit (BIG-5) und einer Schätzung der allgemeinen kognitiven Fähigkeiten mittels Schulleistungstests hinaus bezüglich psychischer Gesundheit. Bei Rossen und Kranzler (2009) erklärte der MSCEIT nach Kontrolle von genereller kognitiver Fähigkeit und Persönlichkeit 1-4% der Varianz von Alkoholkonsum und positiven Beziehungen zu anderen; keine inkrementelle Varianzerklärung ergab sich hingegen für akademische Leistungen und psychologisches Wohlbefinden. Auch bei Amelang und Steinmayr (2006) zeigte sich bei verschiedenen Leistungskriterien für Emotionale Intelligenz (TEMINT) keine inkrementelle Validität über Intelligenz und Gewissenhaftigkeit hinaus, ebenso wie Barchard (2003) durch Emotionale Intelligenz (MSCEIT) für akademischen Erfolg über Intelligenz- und Persönlichkeitsfaktoren hinaus. Die Befundlage ist insgesamt noch zu gering und zu heterogen, um eine abschließende Einschätzung bezüglich der Prädiktionskraft von Emotionale-Intelligenz-Tests vorzunehmen.

Allgemeine Bewertung. Wie ist die Validität der Tests einzuschätzen? Bezogen auf die faktorielle Struktur gibt es eine hohe Übereinstimmung bezüglich eines Modells mit drei Faktoren; Zweig 1 und 2 des Mayer'schen Modells lassen sich häufig nicht als separate Faktoren abbilden. Um höhere Konvergenzen zwischen Theorie und Empirie zu erhalten, erscheint es erforderlich, die theoretischen Konzeptionen und/oder die Qualität der Messinstrumente zu überarbeiten.

In den dargestellten Untersuchungen zeigten die Tests überwiegend konvergente Validität und korrelierten mit verwandten Konstrukten, jedoch zum Teil deutlich geringer als erwartet. Zu berücksichtigen ist die teilweise unzureichende psychometrische Qualität eingesetzter Verfahren: Vor allem für Empathie oder andere emotionsbezogene Fähigkeiten gibt es – anders als bei Intelligenz oder zu den BIG-5 – bislang nur wenige etablierte Erfassungsinstrumente und es werden häufig Selbstbeschreibungsverfahren eingesetzt. Heterogen stellt sich die Befundlage bezüglich des Zusammenhangs von Emotionale-Intelligenz-Tests und Intelligenz-Tests dar; hier ist weitere Forschung dringend erforderlich. In den Studien erwiesen die eingesetzten Emotionale-Intelligenz-Tests hinreichende diskriminante Validität zu Persönlichkeitstests. Trotz verschiedener positiver Ergebnisse bleibt die Prädiktionskraft von Emotionaler Intelligenz vor allem in Bezug auf Leistungsindikatoren bislang deutlich hinter den von Goleman formulierten Erwartungen zurück und es wird auch nicht dieselbe Varianzaufklärung wie durch Intelligenz erreicht (Matthews, Zeidner, & Roberts, 2012; Murphy & Sideman, 2006).

Ausgangspunkt für Erkenntnisfortschritte in diesen Bereichen ist die (Weiter-) Entwicklung von Messinstrumenten für fähigkeitsbasierte Emotionale Intelligenz (Van Rooy, Viswesvaran, & Pluta, 2005). Diese Weiterentwicklung sollte nicht nur eine Verbesserung vorhandener Verfahren beinhalten, sondern auch alternative Ansätze in den Fokus nehmen.

Tabelle 1: Tabellarische Übersicht über Korrelationen der zitierten Studien

Autoren	Emotionale Intelligenz und verwandte Konstrukte	Emotionale Intelligenz und Intelligenz	Emotionale Intelligenz und Persönlichkeit
Austin (2010)	STEM – STEU $r=0.29^{**}$ STEM – MSCEIT $r=0.36^{**}$ STEU – MSCEIT $r=0.33^{**}$ STEU – „Erkennen von Emotionen auf Gesichtern“ $r=0.30^{**}$ STEU – „Reaktionszeit Identifikation von Trauer“ $r=0.21^{**}$ STEM – „Erkennen von Emotionen auf Gesichtern“ $r=0.07$ STEM – „Reaktionszeit Identifikation von Trauer“ $r=0.13$	STEM – Verbale Intelligenz (Wortschatztest) $r=0.12$ STEM – Intelligenz (Buchstabenreihen) $r=0.07$ STEU – Verbale Intelligenz (Wortschatztest) $r=0.32^{**}$ STEU – Intelligenz (Buchstabenreihen) $r=0.10$ MSCEIT – Verbale Intelligenz (Wortschatztest) $r=0.02$ MSCEIT – Intelligenz (Buchstabenreihen) $r=-0.06$	
Austin, Farrelly, Black & Moore (2007)			MSCEIT – IPIP (Neurotizismus) $r=-0.07$ MSCEIT – IPIP (Extraversion) $r=0.03$ MSCEIT – IPIP (Offenheit) $r=0.11$ MSCEIT – IPIP (Verträglichkeit) $r=0.21^{**}$ MSCEIT – IPIP (Gewissenhaftigkeit) $r=-0.01$
Blickle et al. (2009)	TEMINT ¹ – Empathie $r=-0.26^{**}$ TEMINT ¹ – DANVA $r=0.26^{**}$		TEMINT ¹ – NEO-FFI (Neurotizismus) $r=0.04$ TEMINT ¹ – NEO-FFI (Extraversion) $r=-0.11$ TEMINT ¹ – NEO-FFI (Offenheit) $r=-0.21^{**}$ TEMINT ¹ – NEO-FFI (Verträglichkeit) $r=-0.13$ TEMINT ¹ – NEO-FFI (Gewissenhaftigkeit) $r=-0.08$
Blickle et al. (2011)	TEMINT ¹ – MSCEIT (gesamt) $r=-0.22^{**}$ TEMINT ¹ – MSCEIT (Emotionsverständnis) $r=-0.34^{**}$		
Blickle, Kramer & Mierke (2010)		TEMINT ¹ – IST (verbale Skalen) $r=0.02$ TEMINT ¹ – WPT $r=0.10$	
Choi & Kluemper (2012)	TEMINT ² – MSCEIT (gesamt) $r=0.10$ TEMINT ² – SREIS $r=0.21^*$ TEMINT ² – OREIS $r=0.02$		
Ciarrochi, Caputi & Mayer (2003)	MEIS (Gesamt) – LEAS $r=0.15$ MEIS (Emotionswahrnehmung) – LEAS $r=0.20^*$ MEIS (Beziehungen) – LEAS $r=0.21^*$		
Farrelly & Austin (2011)	MSCEIT – IPT-15 $r=0.29^{**}$ (Studie 1) bzw. $r=0.07$ (Studie 2) MSCEIT – „Reaktionszeit Identifikation von Trauer“ $r=-0.01$ (Studie 1) bzw. $r=0.23^{**}$ (Studie 2) MSCEIT – „Reaktionszeit Identifikation von Freude“ $r=0.12$ (Studie 1)	MSCEIT – Gf/Gc QTB „gf“ $r=0.04$ (Studie 1) MSCEIT – Gf/Gc QTB „gc“ $r=0.03$ (Studie 1) bzw. $r=0.24^{**}$ (Studie 2) MSCEIT – Raven's Progressive Matrices $r=0.09$ (Studie 2)	
Mayer, Caruso & Salovey (2000)	MEIS – Empathie $r=0.33^{**}$	MEIS – Verbale Intelligenz $r=0.36^{**}$	

Tabelle wird fortgesetzt auf S. 24

Fortsetzung Tabelle 1: *Tabellarische Übersicht über Korrelationen der zitierten Studien*

Autoren	Emotionale Intelligenz und verwandte Konstrukte	Emotionale Intelligenz und Intelligenz	Emotionale Intelligenz und Persönlichkeit
Libbrecht & Lievens (2012)	STEM – STEU $r=0.24^{**}$ STEM – MSFDE $r=0.18^{**}$ STEU – MSFDE $r=0.22^{**}$ STEM – SREIS (Emotionen wahrnehmen) $r=0.07$ STEM – SREIS (Emotionen nutzen) $r=0.10^*$ STEM – SREIS (Emotionen verstehen) $r=0.08^*$ STEM – SREIS (Emotionen managen) $r=0.06$ STEU – SREIS (Emotionen wahrnehmen) $r=0.00$ STEU – SREIS (Emotionen nutzen) $r=0.09^*$ STEU – SREIS (Emotionen verstehen) $r=0.08^*$ STEU – SREIS (Emotionen managen) $r=0.03$	STEM – fluide Intelligenz $r=0.05$ STEU – fluide Intelligenz $r=0.10^*$	STEM – IPIP (Neurotizismus) $r=0.02$ STEM – IPIP (Extraversion) $r=0.00$ STEM – IPIP (Offenheit) $r=0.05$ STEM – IPIP (Verträglichkeit) $r=0.20^{**}$ STEM – IPIP (Gewissenhaftigkeit) $r=0.08^*$ STEU – IPIP (Neurotizismus) $r=0.00$ STEU – IPIP (Extraversion) $r=0.02$ STEU – IPIP (Offenheit) $r=0.12^{**}$ STEU – IPIP (Verträglichkeit) $r=0.06$ STEU – IPIP (Gewissenhaftigkeit) $r=-0.01$
MacCann & Roberts (2008)	STEM – STEU $r=0.70^{**}$ STEM – MEIS (Geschichten) $r=0.18$ STEU – MEIS (Geschichten) $r=0.41^{**}$ STEM – TAS-20 (Identifikation von Gefühlen) $r=0.01$ STEM – TAS-20 (Beschreibung von Gefühlen) $r=-0.07$ STEM – TAS-20 (External orientiertes Denken) $r=-0.43^{**}$ STEU – TAS-20 (Identifikation von Gefühlen) $r=-0.12$ STEU – TAS-20 (Beschreibung von Gefühlen) $r=-0.12$ STEU – TAS-20 (External orientiertes Denken) $r=-0.38^{**}$	STEM – Verbale Intelligenz (Wortschatztest) $r=0.41^{**}$ STEU – Verbale Intelligenz (Wortschatztest) $r=0.49^{**}$	STEM – OCEANIC-20 (Neurotizismus) $r=0.02$ STEM – OCEANIC-20 (Extraversion) $r=0.09$ STEM – OCEANIC-20 (Offenheit) $r=-0.11$ STEM – OCEANIC-20 (Verträglichkeit) $r=0.24^*$ STEM – OCEANIC-20 (Gewissenhaftigkeit) $r=-0.05$ STEU – OCEANIC-20 (Neurotizismus) $r=-0.04$ STEU – OCEANIC-20 (Extraversion) $r=-0.06$ STEU – OCEANIC-20 (Offenheit) $r=0.11$ STEU – OCEANIC-20 (Verträglichkeit) $r=0.16^*$ STEU – OCEANIC-20 (Gewissenhaftigkeit) $r=-0.02$
Schulte, Ree, & Carretta (2004)		MSCEIT – WPT $r=0.45^{**}$	MSCEIT – NEO-PI-R (Neurotizismus) $r=-0.28^*$ MSCEIT – NEO-PI-R (Extraversion) $r=0.18^*$ MSCEIT – NEO-PI-R (Offenheit) $r=0.27^*$ MSCEIT – NEO-PI-R (Verträglichkeit) $r=0.27^*$ MSCEIT – NEO-PI-R (Gewissenhaftigkeit) $r=0.23^*$

Anmerkungen:

Die Informationen in dieser Tabelle sind aus Gründen der Übersichtlichkeit knapp gehalten. Für weiterführende Informationen sei auf die Originalquellen sowie das Abkürzungsverzeichnis verwiesen.

DANVA= Diagnostic Analysis of nonverbal Accuracy (Nowicki & Duke, 1994); Gf/Gc QTB= Gf/Gc Quickie Test Battery (Stankov, 1997); IPIP= International Personality Item Pool (Goldberg, et al., 2006); IST-2000-R= Intelligenz-Struktur-Test 2000-R (Amthauer, Brocke, Liepmann, & Beauducel, 2001); LEAS= Level of Emotional Awareness Scale (Lane, Quinlan, Schwartz, Walker, & Zeitlin, 1990); MSFDE= Montreal Set of Facial Displays of Emotion (Beaupré, Cheung, & Hess, 2000); NEO-FFI= NEO-Fünf-Faktoren Inventar (Borkenau & Ostendorf, 1993); NEO-PI-R= Revised NEO Personality Inventory (Costa & McCrae, 1992); OCEANIC-20= Openness conscientiousness extraversion agreeableness neuroticism index condensed (Roberts R. D., 2000); OREIS= other-report EI-scale (Choi & Kluemper, 2012); SREIS= self-report EI scale (Brackett, Rivers, Lerner, Salovey, & Shiffman, 2006); STEU= Situational Test of Emotional Understanding (MacCann & Roberts, 2008); TAS-20= Toronto Alexithymia Scale (Bagby, Parker, & Taylor, 1994); TEMINT= Test zur Erfassung von Emotionaler Intelligenz (Schmidt-Atzert & Bühner, 2002); WPT= Wonderlic Personnel Test (Wonderlic Personnel Test Inc., 2000).

¹ TEMINT= inverses Scoring über Abweichungswerte; d.h. hohe Testwerte bedeuten eine schlechtere Testleistung als geringe Testwerte.

² TEMINT= Abweichungswerte wurden invertiert; d.h. hohe Testwerte bedeuten eine bessere Testleistung als geringe Testwerte.

1.5 Der ViTEIP als neuer Leistungstest zur Erfassung von Emotionaler Intelligenz

Im Rahmen dieser Dissertation kommt ein neuer Leistungstest zur Erfassung von Emotionaler Intelligenz zum Einsatz, der „Videobasierte Test zur Erfassung der Emotionalen Intelligenz im Pflegeberuf“ (ViTEIP). Die Vorgehensweise zur Entwicklung des Tests sowie die finale Version des Tests (ViTEIP-V2) werden in diesem Kapitel beschrieben. Eine sehr umfassende Beschreibung der Testentwicklung findet sich in den Arbeiten von Holtus (2008) und Thomsen (2009).

1.5.1 Theoretische Basis und a priori Annahmen zur Testkonstruktion

Die Testentwicklung des ViTEIP fußt auf nachfolgender theoretischer Basis von Emotionaler Intelligenz (Definition des Konstrukts und Messgegenstand) und a priori Annahmen zum Geltungsgegenstand, der verwendeten Datenquelle, den Scoringmethoden sowie dem Präsentationsmodus.

Definition des Konstrukts und Messgegenstand. Emotionale Intelligenz wird in Anlehnung an die Konzeption von Mayer et al. (Mayer & Salovey, 1993; Salovey & Mayer, 1990) als ein Teilbereich von Intelligenz verstanden, der sich in ein Intelligenzmodell mit „g“ an der Spitze einordnen lässt (MacCann, Joseph, Newman, & Roberts, 2014). Für den „Videobasierten Test zur Erfassung der Emotionalen Intelligenz im Pflegeberuf“ (ViTEIP) wird mit der Fähigkeit „Erkennen von Emotionen Anderer anhand situativer Auslöser“ – im Vergleich zum Emotionale-Intelligenz-Modell von Mayer und Salovey (1997) mit 16 verschiedenen Facetten – ein relativ kleiner und spezifischer Teilbereich als Messgegenstand ausgewählt. Damit soll die Gefahr reduziert werden, dass bei der Entwicklung oder Überprüfung eines neuen Instrumentes zu einem noch relativ jungen Konstrukt wie Emotionale Intelligenz, durch Aggregation verschiedener Fähigkeiten auf Subtestebene, methodische Effekte von Konstrukteffekten nicht mehr klar differenziert werden können (MacCann & Roberts, 2008). Locke (2005) kritisierte, dass Emotionale Intelligenz mittlerweile so weit gefasst würde und so unterschiedliche Komponenten enthalte, dass es kein einleuchtendes Konzept mehr sei. Auch Matthews et al. (Matthews, Zeidner, & Roberts, 2012, p. 119) schlagen vor, Emotionale Intelligenz nicht als ein Konglomerat verschiedener emotionaler Fähigkeiten zu sehen, sondern eine Differenzierung vorzunehmen: *„...Rather than lumping these various qualities¹⁰ together as a unitary EI, our recommendation is that each is studied separately, generating its own set of measures and theories. Better conceptualization and measurement within each domain will then support investigations of interrelationships between them...”*.

¹⁰ Mit „various qualities“ beziehen sich Matthews et al. (Matthews, Zeidner, & Roberts, 2012) auf Temperament, Informationsverarbeitung, Emotionsregulation, Emotionales Wissen und Fertigkeiten. (Anmerkung der Autorin)

Es geht im ViTEIP um das Erkennen von Emotionen nach der Beschreibung einer spezifischen Situation. Diese Situationsvorgaben stellen nach Schmidt-Atzert (2003) „situativer Auslöser“ dar. Auch Sharma et al. (Sharma, Gangopadhyay, Austin, & Mandal, 2013, p. 65) verweisen im Zusammenhang ihres „Situational Judgment Tests zur Erfassung von Emotionaler Intelligenz“ auf die Bedeutung des Kontextes, in dem eine Emotion auftritt. Schmidt-Atzert und Bühner (2002) gehen davon aus, dass anhand von Mimik und anderen Ausdruckserscheinungen nur schwer feststellbar ist, wie sich jemand fühlt; emotionale Ausdrücke dienen nach Fridlund (1991) auch der Regulation der sozialen Interaktion und spiegeln nicht automatisch die wahren Gefühle einer Person wider. So zeigte sich beispielsweise bei DeBusk und Austin (2011) kein bedeutsamer Zusammenhang zwischen Emotionaler Intelligenz und der korrekten Identifikation von Gesichtsausdrücken. Informativer als die Mimik und Gestik sind hingegen die „situativen Auslöser“, da bestimmte Ereignisse zu bestimmten Emotionen führten. *„Das emotionale Befinden anderer Menschen lässt sich am besten vorhersagen, wenn man die situativen Umstände kennt, denen jemand ausgesetzt ist“* schreibt Schmidt-Atzert (2003, S. 25). Der ViTEIP soll auf Informationen basieren, die sich auf die situativen Begebenheiten beziehen, in denen sich die Zielpersonen befinden. Über die Kenntnis der Situation schlussfolgern die Probanden, welche Emotionen in welcher Intensität bei den Zielpersonen vorhanden sind.

Geltungsbereich. Zu den Qualitätsanforderungen an einen psychologischen Test gehört nach Moosbrugger und Kelava (2008) eine klare Definition des Geltungsbereichs und somit die Antwort auf die Frage nach dem Personenkreis, für den das Verfahren zum Einsatz kommen soll. Durch eine Spezifizierung des Geltungsbereichs kann einerseits eine höhere Validität und gleichzeitig eine höhere Akzeptanz des Verfahrens für die Anwendung im Alltag erzielt werden. Der ViTEIP wird für die Zielgruppe „Mitarbeiter in Pflege und Betreuung“ konzipiert. Die Autorin nimmt an, dass gerade für Berufe im Gesundheitswesen das „Erkennen von Emotionen Anderer anhand situativer Auslöser“ eine Schlüsselfähigkeit darstellt: Pflegekräfte sind in der Regel nicht anwesend, wenn ihre Patienten emotional bewegende Situationen erleben. Sie treten im Nachhinein in Kontakt mit Patienten, bei denen die Emotionen der Situation „nachwirken“ bzw. Patienten berichten den Pflegenden über die vorgefallene Situation. In beiden Fällen ist es für Pflegekräfte von Vorteil, wenn sie die Emotionen der Betroffenen kognitiv nachvollziehen können. In der Metaanalyse von Joseph und Newman (2010) zeigte sich, dass der Zusammenhang zwischen Facetten Emotionaler Intelligenz und Berufserfolg für Berufsbilder mit höheren Anforderungen an Emotionsregulation (z.B. Verkäufer, Makler) enger war, als bei solchen mit geringeren Anforderungen (z.B. Fabrikarbeiter, Militärpolizist). Auch Roberts et al. (2006) mutmaßen, dass Emotionale Intelligenz nur in bestimmten Berufen ein bedeutsamer Prädiktor von Berufserfolg sei und nennen als Beispiel das

Gesundheitswesen und den Dienstleistungsbereich. Libbrecht et al. (Libbrecht, Lievens, Carette, & Côté, 2014) und Lievens (2013) berichten, dass im Rahmen medizinischer oder pflegerischer Ausbildungen neben der fachlichen Qualifikation vermehrt auf „weiche“ Kompetenzen (soft-skills) wie die Interaktion mit dem Patienten oder effektive Kommunikation Wert gelegt wird. Beauvais et al. (Beauvais, Brady, O'Shea, & Griffin, 2011) untersuchten den Zusammenhang von Emotionaler Intelligenz und pflegerischer Leistung von Krankenschwestern. Vier Skalen des „Six Dimension Scale of Nursing Performance“ korrelierten signifikant mit dem Testergebnis des MSCEIT, woraus die Autoren schlussfolgern, dass das Training von Emotionaler Intelligenz ein Bestandteil der Pflegeausbildung darstellen sollte. Vor dem beschriebenen Hintergrund ist es sinnvoll, einen Test für den genannten Geltungsbereich zu entwickeln.

Datenquelle und Scoring. Aus der Annahme, dass es sich bei Emotionaler Intelligenz um einen Teilbereich von Intelligenz handelt, resultiert die Entscheidung, den ViTEIP in Form eines Leistungstests umzusetzen, bei dem anhand eines objektiven Vergleichswerts eine Einschätzung über richtige und falsche Antworten möglich ist. Die Verwendung des Konsensscorings im Zusammenhang mit der Erfassung von Emotionaler Intelligenz ist mit einiger Kritik verbunden (siehe Kapitel 1.2.1), die nach Ansicht der Autorin nicht zu vernachlässigen ist. Vor allem die Vorgehensweise, die Korrektheit einer Antwort über die Übereinstimmung mit einer Mehrheitsmeinung zu erfassen, widerspricht der grundlegenden Logik der Intelligenzdiagnostik. Bei schwierigen Items, die nur von einer Minderheit der Probanden gelöst werden können, kann gerade die Abweichung von der Norm (d.h. der Mehrheit) intelligentes Verhalten darstellen. Diesem Aspekt wird bei der Verwendung des Konsensscorings keine Rechnung getragen.

Das Targetscoring hingegen legt unabhängig von der Mehrheitsmeinung oder Zusammensetzung einer Normstichprobe fest, dass die Übereinstimmung mit der Angabe der Originalperson als Maßstab für die Richtigkeit einer Probandenantwort gilt. Eine Übereinstimmung mit den Antworten der Originalperson wird immer mit einem höheren Testkennwert gewertet als eine Abweichung von diesen Angaben – unabhängig von den Antworten einer Normstichprobe. Die Nachteile, die das Targetscoring beinhaltet, beziehen sich in erster Linie auf den Erhebungsprozess der Angaben der Originalpersonen. Diese können durch eine sorgfältige Vorgehensweise bei der Datenerhebung minimiert werden und sind somit aus Sicht der Autorin handhabbar. Um unterschiedliche Scorings vergleichen zu können, soll der ViTEIP jedoch mittels Target- und Konsensscoring auswertbar sein.

Präsentationsmodus. Roberts et al. (Roberts R. D., et al., 2006) empfehlen, bei der Entwicklung von Emotionale-Intelligenz-Tests auch auf multimediale Möglichkeiten zurückzugreifen. Neben computerbasierten Tests, adaptiven Verfahren und computerunterstützten Testauswertungen, ist

es heute technisch unproblematisch, Tests oder Testteile in anderer Form als rein schriftlich darzubieten. Gerade im Bereich der Sozialen- und Emotionalen-Intelligenz-Diagnostik erscheint es sinnvoll, die Möglichkeiten videounterstützter Testungen zu nutzen, da dieses Präsentationsformat verschiedene Vorzüge verspricht. Lievens und Sackett (2006) sprechen Videos eine höhere Realitätsnähe zu, was dazu führt, dass Probanden die für die jeweilige Situation geeignete Reaktion vermehrt zeigen. Zudem könne durch die hohe Wiedergabetreue der Videopräsentation die Übereinstimmung mit den Kriterien verbessert werden, was wiederum zu einer höheren prädiktiven Validität führt (Christian, Edwards, & Bradley, 2010). Schlussendlich gehen Lievens und Sackett davon aus, dass videobasierte Tests eine geringere Korrelation zu kognitiven Fähigkeiten aufweisen und damit nicht von anderen Fähigkeiten wie bspw. der Lesefähigkeit abhängen. Kanning et al. (Kanning, Grewe, Hollenberg, & Hadouch, 2006) konnten zeigen, dass videobasierte Testverfahren weniger Leseverständnis erfordern als schriftlich dargebotene Items.

Zur Erhöhung der Nachvollziehbarkeit und Realitätsnähe werden im ViTEIP den Probanden Videosequenzen präsentiert, in denen die beschriebene Person sowie die Räumlichkeiten, in denen das beschriebene Erlebnis stattgefunden hat, zu sehen sind. Die Videos werden von einer mündlichen Schilderung der Situation begleitet; zu beachten ist, dass die mündlichen Informationen nicht identisch zu den visuellen Eindrücken sind. Durch die Videos wird angestrebt, den Probanden zusätzlich zu der verbalen Beschreibung einen Eindruck des Gesamtkontexts zu bieten. Die Kenntnis über situativen Rahmenbedingungen, die zur Auslösung der Emotion auf Seite der Originalperson geführt haben (Siemer, Mauss, & Gross, 2007), soll den Probanden die Einschätzung erleichtern.

1.5.2 Testentwicklung ViTEIP-V1

Nachfolgend werden die Begriffe „Item“ und „Situation“ verwendet: Item bezieht sich auf eine einzelne Emotion und die dazu gehörige Einschätzung der Intensität. Situationen bestehen aus mehreren Items und bilden inhaltlich zusammengehörige Cluster von Emotionen, so wie sie von den Originalpersonen als Ereignis geschildert wurden.

Erhebung der Erlebnisbeschreibungen mittels Interviews. Um das Targetscoring (Beschreibung siehe Kapitel 1.2.1) verwenden zu können, wurden in einer Altenhilfeeinrichtung und einer Rehaklinik Mitarbeiter, Bewohner und Patienten (=Originalpersonen) nach kritischen Situationen im Rahmen von Interviews befragt (vgl. Critical Incident Technique nach Flanagan, 1954). Kritische Situationen waren definiert als *„Situationen, die nicht länger als 14 Tage in der Vergangenheit lagen, in denen auf Seite des Patienten ein oder mehrere Gefühle auftraten und die typisch und relevant für den Alltag in der entsprechenden Einrichtung sind“*. Die Interviews führte eine in Gesprächsführung geschulte Psychologiestudentin. Neben der Schilderung der Situation und den

Beteiligten sowie einigen Angaben zur eigenen Person, wurden die Originalperson nach dem Vorhandensein und Ausprägungsgrad verschiedener Emotionen befragt. Die Interviews wurden während des Gesprächs protokolliert; die dokumentierten Situationen und Emotionen bilden die Grundlage der Testitems. Alle Originalpersonen erhielten zur Wahrung der Anonymität fiktive Nachnamen.

Stichproben. Es wurden mit 34 Mitarbeitern und 18 Patienten Gespräche zu kritischen Situationen geführt. Es wurde darauf geachtet, dass die Patienten/Bewohner nicht an einer diagnostizierten dementiellen Erkrankung litten. Insgesamt 17 Situationsbeschreibungen von elf Patienten/Bewohnern (neun Frauen, zwei Männer) konnten letztendlich erhoben werden. Das Alter der Stichprobe reicht von 67-86 Jahren mit einem Mittelwert von 80.6 Jahren in der Altenhilfe und 75.5 Jahren in der Rehaklinik. Genaue Angaben zu Geschlecht, Alter und Aufenthaltsdauer in den Institutionen finden sich in *Tabelle 2*. Zusätzlich zu den Erlebnissen der Patienten wurden zwei Situationsbeschreibungen ausgewählt, die sich zwischen Mitarbeitenden der Altenhilfeeinrichtung ereignet hatten. Die beiden Mitarbeiterinnen im Alter von 27 bzw. 26 Jahren verfügen über mehrjährige Berufserfahrung im Pflegebereich.

Tabelle 2: Stichprobenbeschreibung der Originalpersonen

		Rehaklinik	Altenhilfeeinrichtung (Bewohner)	Altenhilfeeinrichtung (Mitarbeiter)
Geschlecht	Anzahl Frauen	4	5	2
	Anzahl Männer	2	0	0
Alter (in Jahren)	Spannweite	67-87	78-83	27-36
	Mittelwert	75.5	80.6	31.5
Bisherige Verweildauer am Datum des Interviews	< 2 Tage	2		
	3-7 Tage	2		
	1-8 Wochen	2		
	2-6 Monate		1	
	7-12 Monate		1	
	1-3 Jahre		2	
	> 3 Jahre		1	

Entwicklung der Situationsbeschreibungen und Erstellung der Videos. Die Entwicklung der Situationsbeschreibungen für den Test erfolgte über ein mehrstufiges Vorgehen, bei dem zunächst zwei Psychologie-Diplomandinnen aus den Gesprächsprotokollen der 19 Situationen separat voneinander je eine Textversion formulierten. Diese Versionen sollten folgenden Kriterien genügen:

- Es handelt sich um ein klar abgrenzbares Ereignis.
- Die Erlebnisbeschreibung beinhalten Informationen zur Person, zum Hintergrund der Situation und zur konkreten Situation.

- Die Situation ist auch ohne pflegebezogenes Fachwissen und ohne spezifische Kenntnis über die Einrichtung nachvollziehbar.
- Aus der Beschreibung ist eine Einschätzung der emotionalen Reaktion möglich, ohne dass in der Formulierung Hinweise auf Emotionen gegeben werden.
- Die Beschreibung enthält ca. 150-200 Wörter.

Beide Formulierungen der Situationen wurden zwei erfahrenen Pflegekräften vorgelegt, mit der Bitte, die Situationsbeschreibungen in Bezug auf Verständlichkeit und die Realitätsnähe im Pflege-Alltag zu beurteilen. Aufgrund der Rückmeldungen der Pflege-Experten wurde die als verständlicher beurteilte Beschreibung ausgewählt, kleinere Formulierungsänderungen vorgenommen und eine Situation aufgrund einer zu hohen Spezifität für die Einrichtung ausgeschlossen. Die verbliebenen 18 Situationsbeschreibungen wurden anschließend von einer Sprecherin mit ruhigem, sachlichen Ton und deutlicher Aussprache auf Tonband aufgenommen.



Anmerkungen:

Schriftliche Einverständniserklärungen der Originalpersonen zur Verwendung der Videos liegen vor.

¹ Darstellung der betroffenen Person bei einer neutralen, nicht-emotionalen Tätigkeit

² Visueller Hinweis auf die konkrete Situationsbeschreibung

Abbildung 7: Ablauf der Videosequenzen am Beispiel der Situation 7 „Frau Hoffmann bestellt Inkontinenzeinlagen“

Für jede Situation wurde mit schriftlicher Zustimmung der Betroffenen, ein kurzer Videofilm gedreht. Die Videos wurden so geschnitten, dass sie zu den aufgenommenen verbalen Situationsbeschreibungen passten. In den Videosequenzen wird zunächst die Originalperson näher vorgestellt (Alter der Person, derzeitige Lebenssituation, Gesundheitszustand, Grund und Dauer des Aufenthalts in der angesprochenen Einrichtung usw.), anschließend wird die real erlebte Situation von einer Sprecherin geschildert. Während der Vorstellung und Schilderung ist die Originalperson in ihrem Umfeld (Altenheim oder Rehaklinik) bei einer neutralen, nicht-emotionalen Tätigkeit zu

sehen. Mit einem visuellen Signal (siehe *Abbildung 7*, rechtes Bild) wird der Beginn der eigentlichen Situationsbeschreibung verdeutlicht.

Entwicklung des Testhefts. Es wurde ein Testheft erstellt, in dem die Probanden durch Ankreuzen einschätzen, welche Emotionen in welcher Intensität (vierstufige Skala von „nicht vorhanden“ bis „stark vorhanden“) bei den Personen der präsentierten Situationen vorliegen. Die Liste der Gefühle wurde in Anlehnung an den EMO-16 (Schmidt-Atzert & Hüppe, 1996) entwickelt; Langeweile, Sehnsucht, Mitgefühl und sexuelle Erregung wurden aus Gründen des fehlenden Bezugs zu den Situationen ausgeschlossen, dafür wurden Enttäuschung, Dankbarkeit und Ekel in die Aufzählung aufgenommen. Emotionen, die in der jeweiligen Situation offensichtlich keine Rolle spielen, wurden blass gedruckt und müssen von den Probanden nicht bearbeitet werden. Dieses Vorgehen wurde gewählt, da sich den Untersuchungen von Witten (2003) und Mertin (2004) gezeigt hatte, dass eine zu hohe Anzahl sogenannter Null-Items (d.h. Items, in denen die Angabe der Originalperson „Emotion nicht vorhanden“ lautet) sich negativ auf die Testgüte auswirkt.

Durchführung einer Probetestung. Die Version ViTEIP-V1 beinhaltet 18 Situationen mit insgesamt 184 Items. Vor der eigentlichen Untersuchung wurde eine Probetestung mit insgesamt sechs Psychologiestudenten durchgeführt, um Verständlichkeit der Instruktion, die Funktionstüchtigkeit der Technik und benötigte Testdauer zu prüfen.

Einsatz des ViTEIP-V1. Die Stichprobe 1 (Beschreibung siehe 3.1) bearbeitete den ViTEIP-V1. Für den ViTEIP-V1 wurden zufriedenstellende Reliabilitäten von $\alpha=0.82$ erreicht. Auch die Korrelationen zu den anderen eingesetzten Verfahren fielen im Großen und Ganzen erwartungsgemäß aus. Eine detaillierte und ausführliche Darstellung der Ergebnisse dieser Testversion findet sich bei Holtus (2008) und Thomsen (2009). Mit 18 Situationen und einer Bearbeitungszeit von rund 60 Minuten ist der ViTEIP-V1 jedoch deutlich zu lang und den Rückmeldungen der Probanden zufolge auch zu anstrengend und langwierig in der Bearbeitung¹¹.

1.5.3 Endversion ViTEIP-V2

Da die Version-V1 aufgrund der langen Bearbeitungsdauer nicht zum Einsatz für diagnostische Fragestellungen geeignet ist, wurden die kürzeren ViTEIP-Versionen V2a und V2b entwickelt.

Itemselektion. Zunächst wurden 3 Situationen ausgeschlossen, die eine sehr geringe bzw. negative Trennschärfe ≤ 0.20 aufwiesen (Situation 8, 10, 17). Da Situation 18 nunmehr die einzige Schilderung eines Mitarbeitererlebnisses war, wurde diese aus inhaltlichen Gründen ausgeschlossen. Für

¹¹ Marentette, Meyers, Hurtz und Kuang (2012) warnen vor konstrukt-irrelevanten Methodeneffekten, die mit zunehmender Testlänge auftreten können.

Situation 14 lag nur ein Video ohne Person vor und wurde daraufhin eliminiert. Da es im ViTEIP in erster Linie um die Fähigkeit geht, vorhandene Emotionen zu erkennen, wurde die Itemselektion auf Items beschränkt, in denen die Originalpersonen die Antwort „Emotion nicht vorhanden“ gegeben hatten. Über den Test hinweg wurden sehr leichte und sehr schwierige Items dieser Kategorie entfernt. Sehr leichte Items waren dadurch gekennzeichnet, dass die Probanden einen mittleren Abweichungswert von den Angaben der Originalpersonen von ≤ 0.50 erzielten, es also eine extrem hohe Übereinstimmung mit den Angaben der Originalperson in Bezug auf das „Nichtvorhandensein“ dieser Emotion gab. Als schwierige Items wurden die Items bezeichnet, bei denen die Abweichung des Mittelwerts zu der Angabe der Originalperson mehr als 2 Standardabweichungen betrug. Nachdem durch die Itemselektion in Situation 3 nur noch die Hälfte der ursprünglichen Items vorhanden war, wurde auch diese Situation aus dem Test entfernt. Es wurde in Kauf genommen, dass zwei Originalpersonen in je 3 Situationen auftreten und die Emotion Neid aus der Liste entfernt werden musste. Das Antwortheft wurde entsprechend angepasst. *Tabelle A-1 (Anhang A)* bietet einen Überblick über die Situationen der Version V2.

Zwei verschiedene Antwortformate. Für die Version V2b fand eine Veränderung des Antwortformats statt: Hier erhalten die Probanden nicht mehr eine Liste von Emotionen, die auf einer Likert-Skala in Bezug auf die Intensität eingeschätzt werden müssen, sondern die Emotionen werden in einem separaten Feld dargeboten. Aufgabe der Probanden ist es, die Emotionen drei Kategorien zuzuordnen („stark vorhandene Emotionen“, „mittel vorhandene Emotionen“ und „schwach vorhandene Emotionen“). In dem ursprünglichen Feld verbleibende Emotionen werden als „nicht vorhanden“ gewertet. Diese Änderungen wurden vorgenommen aufgrund der Rückmeldungen der Probanden, dass die Bearbeitung der Emotionslisten extrem langatmig sei. In *Anhang A, Abbildung A-1 und A-2* findet sich eine exemplarische Darstellung der Antwortformate.

Bewertung des ViTEIP-V2. Mit dem „Videobasierten Test zur Erfassung der Emotionalen Intelligenz im Pflegeberuf“ (ViTEIP) wurde ein Verfahren entwickelt, das die Emotionale-Intelligenz-Facette „Erkennen von Emotionen Anderer anhand situativer Auslöser“ in Form eines Leistungstests erfasst. Es existieren klaren Regeln für die Auswertung der Probandenantworten in Bezug auf die Korrektheit der Antworten; dies gilt sowohl für das Target- als auch Konsensscoring. Da der ViTEIP in der vorliegenden Arbeit zum ersten Mal zum Einsatz kommt, sei für Informationen bezüglich der Testgütekriterien auf die nachfolgenden Kapitel verwiesen.

Wie später beschrieben werden wird, weist der ViTEIP-V2 trotz verschiedener positiver Ergebnisse einzelne Kritikpunkte auf, die einen Einsatz für diagnostische Fragestellungen (z.B. Personalauswahl, Eignungsentscheidungen) aus heutiger Sicht noch nicht empfehlen. Vor allem die Frage nach der

Validität des ViTEIP-V2 bleibt zum Teil unbeantwortet, wobei anhand des vorliegenden Untersuchungsdesigns nicht geklärt werden kann, ob die Unzulänglichkeiten aus dem speziellen Verfahren oder generell aus dem Konstrukt Emotionale Intelligenz herrühren (vgl. Matthews, Emo, Roberts, & Zeidner, 2006, p. 17).

An dieser Stelle soll auf zwei Aspekte verwiesen werden, auf die im Rahmen der Auseinandersetzung mit den Limitationen der vorliegenden Arbeit (Kapitel 5.6) ausführlicher eingegangen werden wird. Zum einen handelt es sich bei der Stichprobe der Originalpersonen in erster Linie um weibliche Personen. Der zweite Aspekt, bezieht sich darauf, dass in den von den Originalpersonen geschilderten Situationen überwiegend negative Emotionen vorherrschen. Damit deckt der ViTEIP nur einen Teilbereich von „Erkennen von Emotionen Anderer anhand situativer Auslöser“ ab; der Umgang mit positiven Emotionen bleibt weitestgehend unberücksichtigt. Sowohl die Stichprobenzusammensetzung als auch die Itemauswahl bezüglich der Emotionen ergibt sich jedoch aus der gewählten Konstruktionsweise des ViTEIP: die Patienten und Bewohner der Gesundheitseinrichtungen, die sich bereit erklärten ihre erlebten Situationen zu schildern, waren überwiegend Frauen bzw. schilderten vor allem Begebenheiten mit negativen Emotionen. Auf ein steuerndes Eingreifen bezüglich einer Gleichverteilung des Geschlechts der Originalpersonen oder der Schilderung positiver Situationen wurde bewusst verzichtet.

2 Forschungsfragen und Hypothesen

Aus den vorangegangenen Beschreibungen wird deutlich, dass trotz intensiver Forschungsbemühungen die Messung von Emotionaler Intelligenz nach wie vor mit verschiedenen Schwierigkeiten assoziiert ist. Daraus ergibt sich die Frage, wie die Validität von Emotionale-Intelligenz-Tests verbessert werden kann. Ein möglicher Ansatzpunkt, um diese Frage zu beantworten liegt sicherlich in einer grundlegenden Auseinandersetzung mit dem Konstrukt Emotionale Intelligenz, wie sie beispielsweise von Antonakis et al. (Antonakis, Ashkanasy, & Dasbourough, 2009) oder Brody (2009) angeregt wird. Die Diskussion, ob es Emotionale Intelligenz tatsächlich gibt und wie ein entsprechendes Konstrukt konzipiert sein könnte oder sollte, wird im Rahmen der vorliegenden Arbeit jedoch nicht vertiefend geführt, sondern nur am Rande im Rahmen der Diskussion gestreift. Der Ansatzpunkt für diese Dissertation ist ein anderer: Wie der Titel der Arbeit „Methodische Fragen bei der Erfassung von Emotionaler Intelligenz“ zeigt, besteht die Annahme, dass verschiedene methodische ‚Stellschrauben‘ existieren, die bei der Erfassung von Emotionaler Intelligenz mittels Leistungstest eine Rolle spielen und somit als Einflussgrößen auf die Validität in Frage kommen. Diese Parameter genauer zu betrachten ist Kern der vorliegenden Arbeit.

Eine wesentliche Stellschraube ist dabei der Auswertungsansatz (Scoring), der die Grundlage liefert, nach welchen Kriterien eine Probandenantwort als richtig oder falsch bewertet wird. Wie unter 1.2.1 beschrieben, ist es nach wie vor unklar, mit welchem Scoring Emotionale-Intelligenz-Tests am besten ausgewertet werden sollen. Daher werden im Rahmen dieser Dissertation unterschiedliche target- und konsensbasierte Scorings untersucht. Ziel ist es, Empfehlungen aussprechen zu können, welche Scorings sich für die Auswertung bei Emotionale-Intelligenz-Tests eignen. Im Laufe der Analysen der vorliegenden Daten zeigte sich, dass auf Seiten der Probanden Antworttendenzen (Bias) auftreten, die die Ergebnisse bei einer targetbasierten Auswertung verzerren; eine entsprechende Erläuterung wird in Kapitel 2.1.1 erfolgen. Um dieser Situation zu begegnen, sollen neue targetbasierte, bias-robuste Scorings, entwickelt und erstmals angewandt werden.

Bei diagnostischen Instrumenten ist es erforderlich, diejenigen Effekte so gering wie möglich zu halten oder zu kontrollieren, die durch das spezielle Verfahren der Datenerhebung und nicht durch das Merkmal selbst bedingt sind. Daher werden neben der Untersuchung der Scorings im Rahmen dieser Arbeit zwei weitere methodische Einflussgrößen in den Blick genommen, die in der Literatur zu Emotionaler Intelligenz nur marginal thematisiert werden, für Testentwicklungen oder die Einordnung eines Test(-ergebnisse)s jedoch von Bedeutung sind: der Präsentationsmodus der Items und das Format, in dem die Antwortoptionen präsentiert werden.

In der vorliegenden Arbeit geht es somit um folgende Forschungsfragen:

- Was passiert, wenn ein Emotionale-Intelligenz-Test mit unterschiedlichen Scorings ausgewertet wird? Welche Scorings sind für die Anwendung zu empfehlen?
- Inwiefern wirkt sich ein geändertes Antwortformat auf die Ergebnisse im ViTEIP-V2 aus?
- Ist ein videobasierter Präsentationsmodus für den ViTEIP-V2 geeignet?

In *Tabelle 3* sind die Hypothesen zu den Forschungsfragen für den Überblick komprimiert dargestellt; eine Herleitung und Erläuterung der Hypothesen erfolgt in den nachfolgenden Unterkapiteln. Die Untersuchung der Hypothesen zu den Scorings erfolgt mittels der Emotionale-Intelligenz-Tests ViTEIP-V2 und TEMINT.

Tabelle 3: Tabellarischer Überblick über die Hypothesen der vorliegenden Arbeit

Forschungsfrage	Nr.	Hypothese¹
Auswertungsstrategien		
Vergleich von target- und konsensbasierten Scorings	1a	Korrelation zwischen Scorings signifikant auf dem Niveau $p < 0.01$
	1b	Korrelation zu anderen Verfahren: Konsens = Target
Neue Scorings mit biasfreien Kennwerten	1c	Antwortbias bei „Profil-Gesamt“ geringer als bei „Target-Übereinstimmung“
	1d	Antwortbias bei „Detektionsleistung“ geringer als bei „Target-Übereinstimmung“
	1e	Bestätigung der konvergenten Validität
Antwortformate	2a	Zentrale Kennwerte: Ankreuzen = Zuordnen
	2b	Korrelationsmuster: Ankreuzen = Zuordnen
Präsentationsmodus	3a	Korrelation mit verbaler Intelligenz: Text > Video
	3b	Akzeptanz des Verfahrens durch die Probanden: Video > Text
	3c	Korrelation mit Merkfähigkeit: Video > Text
	3d	Korrelation mit Aufmerksamkeit: Video > Text

Anmerkungen:

Nr. = Nummerierung der Hypothesen.

¹ Ausführliche Erläuterung der Hypothesen im nachfolgenden Text.

2.1 Auswertungsstrategien

In der vorliegenden Arbeit findet ein Vergleich von target- und konsensbasierten Scorings statt. Ein wesentlicher Unterschied zwischen diesen beiden Ansätzen besteht darin, wem die Kompetenz zugesprochen wird, als Referenz für die Festlegung der Korrektheit der Antwort zu dienen: den Betroffenen (d.h. Originalpersonen) bei der Targetauswertung oder Außenstehenden (d.h. einer Referenzgruppe) bei Konsensauswertung? Aufgrund inhaltlicher Nähe und hoher Konvergenz zwischen Konsens- und Expertenscoring (Legree, Psotka, Tremble, & Bourne, 2005; Mayer J. D., Salovey, Caruso, & Sitarenios, 2003), wird in der vorliegenden Arbeit auf eine Analyse des

letztgenannten verzichtet. Die Forschungsfrage lautet: Was passiert, wenn ein Emotionale-Intelligenz-Test mit unterschiedlichen Scorings ausgewertet wird? Einerseits ist es plausibel, dass sich Ergebnisse verändern, wie anhand des Beispiels eines klassischen Intelligenz-Tests verdeutlicht werden kann. Hier würde man kaum erwarten, dass eine Auswertung mittels Konsensscoring durchgängig zu denselben Ergebnissen führt, wie eine Auswertung anhand objektiver Vorgaben (z.B. mathematischer Grundregeln). Wie Barchard et al. (Barchard, Hensley, & Anderson, 2013) annehmen, wäre eine Übereinstimmung nur für einfache Items, bei denen die Mehrheit der Probanden die richtige Antwort identifizieren kann, zu erwarten.

Andererseits ist aus einer testtheoretischen Perspektive heraus eine Veränderung der Ergebnisse kritisch zu sehen. Die Begründung dafür liegt in der klassischen Testtheorie, wonach sich der beobachtete Messwert einer Person aus Varianzanteilen aus dem Konstrukt und Varianzanteilen aufgrund von Messfehlern zusammensetzt (Moosbrugger H., 2008). Der Messfehler wiederum lässt sich in zwei Komponenten unterteilen: in unsystematische Fehler, die sich über unendlich viele Testwiederholungen ausmitteln (Bühner, 2006) und in systematische Methodeneffekte (Cote & Buckley, 1987; Spector, 2006), zu denen auch Effekte aufgrund von Auswertungsprozeduren zählen. Moosbrugger und Kelava beschreiben „...*Methodeneffekt ist ein Sammelbegriff für verschiedene systematische Varianzquellen, die sich über den Trait hinausgehend auf die Validität der Messung auswirken können...*“ (2008, S. 327). Podsakoff, MacKenzie, Lee und Podsakoff (2003) weisen darauf hin, dass systematische Messfehler problematisch sind, da sie eine alternative Erklärung für beobachtete Zusammenhänge anbieten.

Die Korrelationen zwischen Ergebnissen innerhalb eines Tests bei unterschiedlichen Scorings sollten hoch ausfallen, damit das gleiche Konstrukt repräsentiert wird. Die Anwendung eines anderen Auswertungsprinzips sollte also grundsätzlich keine Auswirkungen auf die Ergebnisse eines Emotionale-Intelligenz-Tests haben und somit nicht die Rangreihe der Probanden in Bezug auf ihre Testkennwerte verändern: Ein Proband A, der in einem Test mit Targetscoring ein höheres Testergebnis erzielt als Proband B, sollte in demselben Test bei Anwendung des Konsensscorings auch ein höheres Ergebnis erhalten als Proband B. Darüber hinaus ist von Interesse, wie die Zusammenhänge zu anderen erfassten Personenmerkmalen ausfallen, wenn andere Scorings zum Einsatz kommen. Bleiben Korrelationen vergleichbar oder ändern sich diese?

Die gleichberechtigte Darstellung der drei verschiedenen Scoringansätze (Konsens-, Experten-, Targetscoring) in der Literatur zu Emotionaler Intelligenz ohne Hinweise auf mögliche Unterschiede in den Ergebnissen, suggeriert dem Leser/Testautor, es stünde ihm grundsätzlich frei, für welches Scoring er sich entscheide. Ob unterschiedliche Auswertungsstrategien tatsächlich zu keinen

bedeutsamen Veränderungen in relevanten Testkennwerten führen, soll mit unten stehenden, im Sinne von Nullhypothesen formulierten Annahmen, geprüft werden.

Korrelative Zusammenhänge. Wenn es sich bei den unterschiedlichen Scorings um vergleichbare Ansätze handelt, äußert sich dies innerhalb eines Tests in einer hohen positiven Korrelation (** $p < 0.01$) der Scorings untereinander (**Hypothese 1a**).

Darüber hinaus wird erwartet, dass unterschiedliche Auswertungsmethoden keine Auswirkungen (d.h. keine signifikanten Unterschiede in der Korrelationshöhe) auf die Korrelationsmuster in Bezug auf die verschiedenen Validitätsaspekte haben (**Hypothese 1b**).

Erläuterung der erwarteten Korrelationsmuster zur Validität. Bezüglich der Validität werden grundsätzlich folgenden Korrelationen erwartet. Die *konvergente Validität* lässt sich durch eine hohe positive Korrelation zwischen Tests, die dasselbe oder eng verwandte Konstrukte (Empathie und Emotionaler Kompetenz) messen, belegen. Es zeigt sich daher eine positive Korrelation zwischen den Emotionale-Intelligenz-Tests VITEIP-V2 und TEMINT. Zudem besteht die Annahme, dass Emotionale Intelligenz eine Intelligenzfacette ist und sich dies in einer moderat positiven Korrelation mit Intelligenztestwerten äußert (vgl. Mayer, Caruso, & Salovey, 2000); in Übereinstimmung mit der Annahme von Mayer, Salovey und Caruso (2008) werden aufgrund der verbalen Testformate höhere Korrelationen zu verbaler als zu figuraler Intelligenz erwartet. Im Sinne der Abgrenzung von bestehenden Konstrukten wird erwartet, dass der VITEIP-V2 und TEMINT *diskriminante Validität* in Bezug auf Persönlichkeit und Merkfähigkeit zeigen. Dies äußert sich in Form von schwachen oder Null-Korrelationen zu den Facetten eines Tests zur Erfassung von Persönlichkeitseigenschaften; die Facetten „Offenheit für Neues“ und „Verträglichkeit“ korrelieren aufgrund des interpersonellen Bezugs der Skalen dabei etwas höher (maximal moderat) mit VITEIP-V2 und TEMINT. Aufgrund des kognitiven Charakters der Emotionale-Intelligenz-Tests können zu Merkfähigkeitstests geringe positive Korrelationen auftreten.

Matthews, Zeidner und Roberts (2012) kritisieren an der aktuellen Forschung zu Emotionaler Intelligenz, dass nach wie vor schlüssige Belege für die Kriteriumsvalidität fehlen. In der zweiten Untersuchung sollten daher anhand von Einschätzungen durch Kollegen und Vorgesetzte verschiedene externe Kriterien und der EKF-Fremd erhoben werden. Trotz gründlicher Vorbereitung und konsequentem Nachhalten war der Rücklauf der Fragebögen so gering, dass Berechnungen bezüglich einer Eignungseinschätzung bzw. der Fremdeinschätzung der Emotionalen Kompetenz nicht sinnvoll waren und entsprechende Hypothesen aus der Arbeit ausgeschlossen wurden.

2.1.1 Entwicklung von Scorings mit biasfreien Kennwerten

Vorliegen eines Antwortbias auf Seiten der Probanden. Ausgangspunkt für die Hypothesen zur Entwicklung von biasfreien Kennwerten waren unerwartete negative Korrelationen zwischen ViTEIP-V2 und TEMINT bei targetbasierten Auswertungen. Als Ursache für diese negativen Korrelationen konnte ein Antwortbias auf Seiten der Probanden identifiziert werden. Nach Bühner (2006) bezeichnet „Antwortbias“ einen systematischen Fehler im Prozess des Messens. Der Antwortbias äußerte sich in der Form, dass Probanden ihre Antworten im ViTEIP-V2 und TEMINT bevorzugt in bestimmten Skalenbereichen wählten und zeigte sich in einer hohen Korrelation in „Häufigkeit der gewählten Antwortoptionen im ViTEIP-V2“ zu „Häufigkeit der gewählten Antwortoptionen im TEMINT“. Exemplarisch bedeutet dies, dass ein Proband A in beiden Tests bevorzugt Antworten im Bereich „Emotion nicht vorhanden“ ankreuzt, ein Proband B hingegen in beiden Tests vor allem Antworten im mittleren Skalenbereich wählt.

Das Vorliegen dieses Antwortbias ist kritisch, da Personen, die ein bestimmtes Antwortverhalten zeigen, bei targetbasierten Auswertungen der Tests systematisch bevorzugt bzw. benachteiligt werden. Die Bevorzugung bzw. Benachteiligung erfolgt jedoch bei den beiden Tests in entgegengesetzter Richtung: im ViTEIP-V2 erhalten Personen, die tendenziell geringe Ausprägungen ankreuzen, bessere Testwerte, im TEMINT erreichen diese Personen eher schlechtere Testwerte. Die umgekehrte Bevorzugung findet sich bei der Personengruppe, die vorzugsweise hohe Ausprägungen ankreuzt. Eine Überprüfung¹² auf das Vorhandensein von testunabhängigen Antworttendenzen ergab, dass die Zusammenhänge zwischen Antwortoptionen bei den anderen eingesetzten Tests mit Likertskalenformat (NEO-FFI, EKF) nur in sehr minimaler Tendenz zu finden waren. Dies legt nahe, dass es sich nicht um ein generelles, testübergreifendes Phänomen der Testbearbeitung handelt, sondern der Bias speziell den ViTEIP-V2 und TEMINT betrifft.

Da an psychologische Tests der Anspruch gestellt wird fair zu sein und keine Bevorzugung oder Benachteiligung bestimmter Personengruppen zuzulassen, werden biasfreie Kennwerte entwickelt. Zwei targetbasierte Scorings, die in der Literatur bislang noch nicht im Zusammenhang des Konstrukts Emotionale Intelligenz genannt wurden, werden konzipiert, angewandt und untersucht.

¹² Pro Person wurden drei neue Kennwerte berechnet, die sich an die von Weijters et al. (Weijters, Geuens, & Schillewaert, 2010) vorgeschlagenen Kodierung zu Antworttendenzen anlehnen. Die Korrelationsmuster der Kennwerte ergeben keinen Hinweis auf das Vorliegen eines testübergreifenden Antwortbias.

Die durchgeführten Analysen sind als Screening zu verstehen, da für die Beurteilung der Existenz eines Antwortbias heterogene Skalen erforderlich sind. Neben heterogenen Items kann der Inhalt durch ausbalancierte Skalen kontrolliert werden (Billiet & McClendon, 2000). Da beim NEO-FFI die Skalen annähernd ausbalanciert sind, könnte bei der Existenz eines „Generalfaktors“ dieser als Antworttendenz interpretiert, als Indikator auf einen Antwortbias verwendet und auf die anderen Verfahren übertragen werden (Bäckström, Björklund, & Larsson, 2009). Die Skalen des NEO-FFI korrelieren nur in erwarteter Richtung miteinander; es gibt in den vorliegenden Daten keinen „Generalfaktor“ der Persönlichkeit, der als Antwortbias interpretiert werden kann.

Ziel ist, Methodiken zur Generierung von Kennwerten zu entwickeln, mit denen der Einfluss der vorhandenen Antworttendenzen der Probanden auf die Gesamtergebnisse signifikant reduziert werden kann.

2.1.1.1 Profilauswertung

Bei der Profilauswertung wird ein Vergleich zwischen den Antwortmustern der Probanden und den Antwortmustern der Originalpersonen vorgenommen. Die Erwartung ist, dass bei einer Auswertung über Profile individuelle Antworttendenzen eine geringere Rolle spielen. So geht es bei der Profilauswertung nicht um die absolute Übereinstimmung zwischen einzelner Angaben der Originalperson und Probandenrating, sondern um das generelle Antwortmuster. Dieses kann insgesamt in Richtung intensivere Gefühle oder schwächere Gefühle verschoben sein und trotzdem zu einem richtig bewerteten Antwortmuster führen.

Die **Hypothese 1c** lautet: Mit der Profilauswertung kann der Einfluss der Antworttendenzen im Vergleich zu Target-Übereinstimmung reduziert werden. Dies zeigt sich in einer signifikant geringeren Korrelation zwischen „Profil-Gesamtwert“ und „Häufigkeit der gewählten Antwortoptionen“ als zwischen „Target-Übereinstimmung“ und „Häufigkeit der gewählten Antwortoptionen“.

Für den Profilvergleich erfolgte eine Auswertung über Situationen¹³, indem das Antwortprofil der Probanden mit dem Antwortprofil der Originalperson verglichen wurde. Als Übereinstimmungsmaß der Profile wurden Spearman-Rangkorrelationen r_s nach *Formel 1* berechnet (Bortz, 2005, S. 233).

$$r_s = 1 - \frac{6 \cdot \sum_{i=1}^n d^2}{n \cdot (n^2 - 1)}, \quad (\text{Formel 1; nach Bortz, 2005})$$

d = Differenz der Rangplätze zwischen Versuchsperson (i) und Angabe der Originalperson

n = Anzahl der Items (d.h. Emotionen) innerhalb der Situation

Um die Differenz d innerhalb einer Situation zu berechnen, wurden den Angaben der Probanden und den Angaben der Probanden innerhalb dieser Situation Ränge zugewiesen. Anschließend wurden die Ränge der Originalperson mit den Rängen eines jeden Probanden verglichen und die jeweilige Differenz zwischen diesen Rängen gebildet. Die quadrierten Differenzen wurden zur Berechnung der Rangkorrelation pro Situation in *Formel 1* eingesetzt. Ein hoher positiver Rangkorrelationswert bedeutet, dass das Antwortmuster eines Probanden eine hohe Ähnlichkeit

¹³ Situationen bestehen aus mehreren Items und bilden inhaltlich zusammengehörige Cluster von Emotionen, so wie sie von den Originalpersonen geschildert wurden.

mit dem Antwortmuster der Originalperson besitzt, dieses Muster sich jedoch in der absoluten Ausprägungshöhe unterscheiden kann. Zur Bildung des „Profil-Gesamtwert“ wurden die so ermittelten Rangkorrelationen für die zwölf Situationen aufsummiert. In *Anhang B, Abbildung A-3* finden sich exemplarische Visualisierungen des Profilvergleichs.

2.1.1.2 Übertragung der Signalentdeckungstheorie auf Emotionale-Intelligenz-Tests

Ein weiteres neues Scoring ergibt sich aus Übertragung des Grundgedankens der Signalentdeckungstheorie (SET) auf Emotionale-Intelligenz-Tests. Die SET stellt ein allgemeines Modell der Entscheidungsbildung dar und ist eine häufig eingesetzte Methode zur Bestimmung der Leistungsfähigkeit von Sinnesorganen unter der Berücksichtigung von Wahrnehmungsprozessen und entscheidungstheoretischen Aspekten (Abdi, 2007; Green & Swets, 1966; Wickens, 2002). Vergleichbar zu Signalentdeckungsaufgaben geht es bei den vorliegenden Emotionale-Intelligenz-Tests für die Probanden um Entscheidungen, ob ein Reiz/Signal (hier: eine bestimmte Emotion in einer gegebenen Situation) vorhanden ist oder nicht (*Tabelle 4*).

Tabelle 4: Übertragung des Vierfelder-Schemas einer Entdeckungsaufgabe auf die Systematik der Emotionale-Intelligenz-Tests bei Targetscoring

		Angabe der Originalperson	
		„Emotion nicht vorhanden“	„Emotion schwach/mittel/stark vorhanden“
Antwort des Probanden	„Emotion nicht vorhanden“	korrekte Verneinung	Auslassung (falsche Zurückweisung)
	„Emotion schwach/mittel/stark vorhanden“	falscher Alarm	richtige Zuordnung (Treffer)

Die SET geht davon aus, dass ein Kontinuum existiert, auf dem die beiden Verteilungen „Rauschen“ (d.h. Angabe der Originalperson= 0) und „Signal“ (d.h. Angabe der Originalperson= 1, 2 (bzw. 3) abgetragen werden können (vgl. *Abbildung 8*). In jedem Durchgang (d.h. bei jedem Item) wird eine unterschiedlich starke Empfindung ausgelöst, woraufhin der Proband entscheiden muss, ob er mit „Signal“ oder „Rauschen“ antwortet. Annahme der SET ist, dass ein Proband gedanklich die Rausch- und die Signalverteilung (Wahrscheinlichkeitsdichten) zueinander ins Verhältnis setzt (d.h. das sogenannte likelihood ratio¹⁴ bildet) und anhand der Lage eines kritischen Wertes entscheidet, ob er mit Signal oder Rauschen antwortet. Welche Antwort ein Proband letzten Endes gibt, hängt dabei nicht nur von der tatsächlichen Erkennensleistung (Kennwert Detektionsleistung \hat{d}') ab, sondern

¹⁴ Verhältnis der Wahrscheinlichkeit, mit der eine bestimmte Beobachtung x aus der Signalverteilung ausgewählt wurde, zu der Wahrscheinlichkeit, dass sie aus der Rauschverteilung ausgewählt wurde. Annahme ist, dass Versuchspersonen zu jeder Beobachtung x den Likelihoodquotienten berechnen und mit „Signal“ antworten, wenn der Quotient größer ist als c .

auch von der individuellen Antworttendenz (Kennwert Reaktionsneigung \hat{c}): Entscheidet sich ein Proband bei Unsicherheit in seiner Antwort eher für „Reiz ist vorhanden“ ($\hat{c} < 0$) oder eher für „kein Reiz vorhanden“ ($\hat{c} > 0$)? Der „ideale“ Beurteiler erreicht einen Wert $\hat{c} = 0$. Mit der Reaktionsneigung bietet die SET damit einen expliziten Kennwert, der auf das Antwortverhalten von Probanden fokussiert und die Detektionsleistung \hat{d}' gilt in der SET als ein valides Maß für die Erkennensleistung, da die Identifikations- und die Fehlalarmrate berücksichtigt werden.

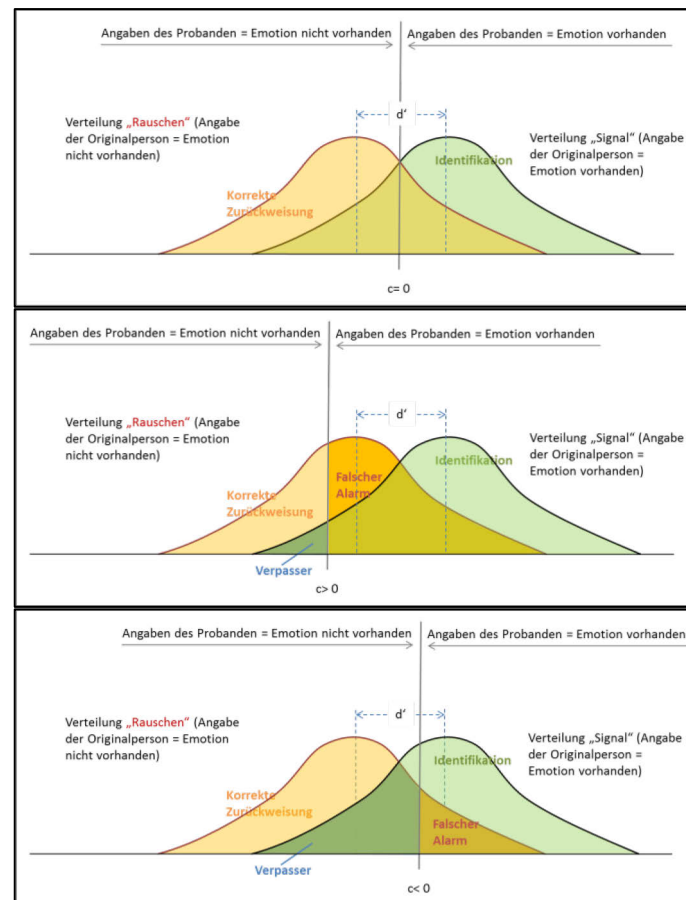


Abbildung 8: Schematische Darstellung der Rausch- und Signalverteilung bei einem Signalentdeckungsexperiment in der Übertragung auf Emotionale-Intelligenz-Tests und verschiedenen Ausprägungen der Lage von \hat{c}

Da es in der vorliegenden Arbeit darum geht, die Grundidee der Signalentdeckung (d.h. die Unterscheidung von Signal- und Rauschdurchgängen) auf Emotionale Intelligenz zu übertragen und nicht darum, ein klassisches Signalentdeckungsexperiment in Detail nachzustellen, werden in der vorliegenden Arbeit zwei Annahmen der SET ohne Prüfung als gegeben vorausgesetzt (Schwaninger, 2005; Wickens, 2002). Zum einen wird davon ausgegangen, dass die Verteilung von „Rauschen“ und „Signal“ jeweils einer Normalverteilung entspricht. Zum anderen ist die Annahme, dass „Rauschen“

und „Signal“ gleiche Verteilungsformen mit gleicher Varianz aufweisen, wobei die Signal-Verteilung auf dem Kontinuum weiter nach rechts verschoben ist. Durch diese Annahmen kann die Detektionsleistung (\hat{d}') als Abstand (in Standardabweichungen) zwischen den beiden Verteilungen im Sinne einer Leistungskomponente und \hat{c} als Modellparameter, der das Entscheidungsverhalten ausdrückt, interpretiert werden. Sollte sich die Signalentdeckungstheorie prinzipiell als Scoring für Emotionale-Intelligenz-Tests eignen, sind Studien über die Gültigkeit der Annahmen erforderlich. Darüber hinaus ist zu beachten, dass sich die Kennwerte der SET klassischerweise auf die Betrachtungsebene des Individuums beziehen und nicht – wie in der vorliegenden Arbeit – auf Werte einer Stichprobe.

Als **Hypothese 1d** wird formuliert: Die Detektionsleistung ist ein geeigneter Indikator, die Leistung des Probanden biasfrei abzubilden. Dies zeigt sich in einer signifikant geringeren Korrelation zwischen der Detektionsleistung und der „Häufigkeit der gewählten Antwortoptionen“ als bei Target-Übereinstimmung.

Zudem korrelieren die Detektionsleistungen in ViTEIP-V2 und TEMINT im Sinne der konvergenten Validität positiv signifikant untereinander sowie mit Intelligenz, Empathie und Emotionaler Kompetenz (**Hypothese 1e**).

2.2 Antwortformate

Was passiert, wenn dasselbe Konstrukt mit denselben Testitems, aber mit einem veränderten Antwortformat erfasst wird? Grundsätzlich ist davon auszugehen, dass – wie im Zusammenhang der Scorings beschrieben wurde – das Antwortformat eines Tests idealerweise zu den unsystematischen Anteilen der Methodenvarianz beiträgt, so dass eine Veränderung des Antwortformats keine Auswirkungen auf die Testkennwerte hat. Nach Podsakoff et al. (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003) können Antwortformate jedoch Einfluss auf das Antwortverhalten von Probanden haben und sind potenzielle Quellen für methodische Fehler. Untersuchungen, die sich im Zusammenhang von Emotionaler Intelligenz mit dieser Frage beschäftigt haben, sind rar. Der Autorin bekannt sind eine Untersuchung von Mayer et al. (2003), in der eine Veränderung von Multiple-Choice zu Ratingskalen im MSCEIT V2.0 zu vergleichbaren Testergebnissen führte und eine Untersuchung von MacCann und Roberts (2008), in der sich mit der Änderung des Antwortformates des STEM auch veränderte Ergebnisse ergaben, beispielsweise eine engere Korrelation zu Intelligenz bei einem Multiple-Choice-Format im Vergleich zu einem Ratingskalenformat.

Inwiefern sich die Ergebnisse im ViTEIP-V2 auf Stichprobenebene verändern, wenn im Testheft ein anderes Antwortformat („Ankreuzen“ vs. „Zuordnen“) als Bearbeitungsgrundlage dargeboten wird, soll mit nachfolgenden Hypothesen getestet werden. Bei „Ankreuzen“ nehmen die Probanden für eine Reihe aufgeführter Emotionen über eine Likertskala eine Einschätzung der Intensität der Emotion vor. Das Antwortformat „Zuordnen“ präsentiert den Probanden alle Emotionen in einem Bearbeitungsfeld. Aus diesem Feld sollen die Emotionen verschiedenen Intensitätskategorien zugeordnet werden; alle Emotionen, die im Bearbeitungsfeld verbleiben, werden der Intensität „Emotion nicht vorhanden“ zugeordnet.

Als **Hypothese 2a** wird im Sinne einer Nullhypothese angenommen: Eine Veränderung von „Ankreuzen“ zu „Zuordnen“ im ViTEIP bringt keine wesentlichen Änderungen in den zentralen Kennwerten (Mittelwert, Varianz, Reliabilität) mit sich.

Die Korrelationsmustern zu anderen Verfahren fallen bei unterschiedlichen Antwortformaten vergleichbar aus (**Hypothese 2b**).

2.3 Präsentationsmodus

Durch die Entwicklungen im Bereich moderner Technologien haben sich in den letzten Jahren Alternativen zu den klassischen Papier-Bleistift-Tests immer mehr verbreitet. Wie in Kapitel 1.5 beschrieben, werden mit videobasierten Tests verschiedene positive Erwartungen verbunden (Chan & Schmitt, 1997; de Meijer, Born, van Zielst, & van der Molen, 2010; Lievens & Sackett, 2006). Es gibt jedoch nicht nur Vorteile in der Nutzung von modernen Medien. Vor allem die Entwicklung eines videobasierten Tests ist mit einem deutlichen Mehraufwand verbunden, der durch den Videodreh, den Videoschnitt etc. entsteht. Auch werden für die Testbearbeitung nicht nur ein Papierausdruck und ein Bleistift wie beim klassischen Paper-Pencil-Test benötigt, sondern weitere technische Utensilien (PC, Beamer, Lautsprecher usw.). Sichergestellt werden muss, dass das intendierte Konstrukt erfasst wird und nicht andere Fähigkeiten, die beispielsweise für die Bedienung eines PCs erforderlich sind oder für die Bearbeitung des Tests eine Rolle spielen.

Ob sich für den ViTEIP der Einsatz einer videobasierten Version anbietet, wird mittels eines Vergleichs zu einer textbasierten Forschungsversion anhand der nachfolgen beschriebenen Hypothesen geprüft.

Korrelation mit verbaler Intelligenz. Um zu verhindern, dass eine Leistung in einem Test vom verbalen Verständnis abhängt, empfehlen MacCann et al. (MacCann, Wang, Matthews, & Roberts, 2010) den Einsatz von Videos. In der Studie von Lievens und Sackett (2006) war die Korrelation mit

kognitiven Fähigkeiten für die videobasierte Version eines Situational Judgment Tests¹⁵ deutlich geringer als dessen schriftlich dargebotene Version. Dieses Ergebnis soll anhand **Hypothese 3a** repliziert werden: Die videobasierte Version des ViTEIP korreliert in signifikant geringerem Ausmaß mit verbaler Intelligenz als die textbasierte Version.

Akzeptanz des Verfahrens durch die Probanden (Soziale Validität). In Chan und Schmitts (1997) Untersuchung erzielte ein videobasierter Test höhere Augenscheinvalidität als die Papier- und Bleistift-Version; zu einem vergleichbaren Schluss kamen auch Richman-Hirsch et al. (Richman-Hirsch, Olson-Buchanan, & Drasgow, 2000). In Anlehnung an diese Ergebnisse wird erwartet, dass die videobasierte Version eine höhere Zustimmung zu den Akzeptanzfragen besitzt als die textbasierte Version (**Hypothese 3b**).

Merkfähigkeit und Aufmerksamkeit. Da beim ViTEIP die Bearbeitung im Antwortheft immer erst nach der Videopräsentation der Situation erfolgt, wird der Einfluss von Merkfähigkeit und Aufmerksamkeit auf die Testleistung untersucht. Bei der Videoversion müssen sich die Probanden die Informationen zur Situation nach einmaligem Sehen der Videos einprägen; die Textversion ermöglicht den Probanden, bei Bedarf die Situationsbeschreibung erneut im Testheft nachzulesen. Die Annahme ist, dass im Vergleich zur textbasierten Version die Korrelation zwischen dem Gesamtergebnis und Merkfähigkeit (**Hypothese 3c**) sowie Aufmerksamkeit (**Hypothese 3d**) bei der videobasierten Version signifikant höher ausfällt.

¹⁵ Eine gute Einführung und Übersicht zu Situational Judgment Tests bieten Weekly & Ployhart (2006) sowie Muck (2013).

3 Methoden

Im Folgenden werden die Stichprobe, die eingesetzten Testverfahren und das Vorgehen zur Auswertung der empirischen Daten genauer beschrieben.

3.1 Stichprobenbeschreibung und Versuchsablauf

Rekrutierung der Versuchspersonen. Die Probanden der Studie 1 wurden über Aushänge, Informationsveranstaltungen in Vorlesungen verschiedener Fachrichtungen und einen Link auf der Homepage der Philipps-Universität Marburg rekrutiert. Als Anreiz zur Teilnahme wurde neben Studiencredits die Möglichkeit zu einer individuellen Rückmeldung über die erreichten Ergebnisse angeboten. Viele psychologische Untersuchungen basieren auf studentischen Stichproben und sind damit nur eingeschränkt verallgemeinerbar (Landy, 2006); die Ergebnisse der ersten Untersuchung sollten daher an einer zweiten, nicht-studentischen Stichprobe repliziert werden. Für die zweite Untersuchung wurde Kontakt zu verschiedenen Altenhilfe- und Krankenpflegeschulen in Hessen aufgenommen. In einem Schreiben wurde über die geplante Untersuchung informiert und um Teilnahme an der Studie gebeten. Alle Schulen erhielten das Angebot, dass die Testleiterin direkt nach der Untersuchung eine kurze Unterrichtseinheit zum Thema Emotionale Intelligenz für die Schüler abhält. Des Weiteren wurde angeboten, für die Schüler eine individuelle Rückmeldung der Testergebnisse zu generieren. In telefonischen Gesprächen mit interessierten Schulleitern wurden alle weiteren Fragen geklärt. Die konkrete Terminabsprache fand anschließend direkt mit den Lehrkräften der teilnehmenden Klassen statt.

Stichprobenbeschreibung. In die Analysen wurden Daten aus zwei Untersuchungen aufgenommen. Aus beiden Stichproben wurden entsprechend der Empfehlung von Maul (2011) für die Analysen die Daten der Teilnehmer ausgeschlossen, bei denen mehr als 10% der ViTEIP- oder TEMINT-Items unbearbeitet waren (Stichprobe 1 $n_{ViTEIP}=2$, $n_{TEMINT}=2$; Stichprobe 2 $n_{ViTEIP}=11$, $n_{TEMINT}=32$). Vor allem in Teilstichprobe 2b fielen vermehrt Probanden durch eine hohe Anzahl an nicht-bearbeiteten TEMINT-Items auf. Möglicherweise ist dies auf ein falsches Verständnis der TEMINT-Instruktion zurückzuführen, da in dieser Teilstichprobe zuvor der ViTEIP-V2b in der Version „Zuordnen“ bearbeitet wurde, wo die Null-Items¹⁶ ignoriert werden sollten. Es wurde analysiert, ob diese Kandidaten bei der Bearbeitung des TEMINT die Instruktion des ViTEIP-V2b „Zuordnen“ übertrugen und daher insgesamt keine Kreuze bei der Emotionsausprägung „nicht vorhanden“ setzten. Diese Annahme konnte jedoch nicht durchgängig bestätigt werden, so dass in diesen Fällen keine

¹⁶ Null-Items= Items, bei denen die Antwort der Originalperson „Emotion nicht vorhanden“ lautet.

Ersetzung der Missings mit „Emotion nicht vorhanden“ vorgenommen, sondern die Datensätze ausgeschlossen wurden. In Stichprobe 1 wurden außerdem Daten von den Personen entfernt, die angaben, Deutsch nicht als Muttersprache erlernt zu haben und gleichzeitig im IST-verbal einen extrem niedrigen Wert erreichten ($n=3$) sowie von einer Person mit Sehbehinderung. In Stichprobe 2 gingen die Ergebnisse von zwei Personen nicht in die Berechnungen ein, die ihre Deutschkenntnisse mit „Grundkenntnisse“ beschrieben.

Wie in *Tabelle 5* aufgeführt, setzte sich die Stichprobe 1 zusammen aus $N=113$ vorwiegend studentischen Teilnehmern (88.5% Studenten; 70.8% weiblich) im Alter von 19 bis 62 Jahre (Mittelwert $M=24.5$, Standardabweichung $SD=8.0$). Selbst nach der oben beschriebenen Daten-selektion erreichten die Stichproben mit $n=138$ (Teilstichprobe 2a) und $n=122$ (Teilstichprobe 2b) noch Umfänge in ausreichend großer Höhe. Die Teilnehmergruppe 2a bestand vorwiegend aus Altenhilfe- und Krankenpflegeschülern (93.5%), weiblichen Geschlechts (79.7%), mit einem Durchschnittsalter von $M=25.0$ Jahren (Altersspanne 17-56 Jahre, $SD=8.7$). Die Stichprobe 2b setzte sich zusammen aus Altenhilfe- und Krankenpflegeschülern (93.4%), weiblichen Geschlechts (73.8%), mit einem durchschnittlichen Alter von $M=25.0$ Jahren (Altersspanne 16-50 Jahre, $SD=8.0$).

Tabelle 5: *Stichprobenbeschreibung*

		Stichprobe 1	Stichprobe 2	
			Teilstichprobe 2a	Teilstichprobe 2b
Anzahl Probanden (N)		113	138	122
Geschlecht	Weiblich	70.8%	79.7%	73.8%
	Männlich	29.2%	20.3%	24.6%
Alter (in Jahren)	Altersspanne	19-62	17-56	16-50
	Mittelwert M	24.5	25.0	25.0
	Standardabweichung SD	8.0	8.7	8.0
Höchster Schulabschluss	Abitur/Fachhochschulabschluss	94.7%	18.9%	17.9%
	Realschulabschluss	1.8%	68.8%	62.5%
	Hauptschulabschluss	0.9%	9.4%	10.7%
Studienfach/Ausbildung	Psychologie	29.2%		
	Studium andere Fachrichtung	59.3%		
	Kein Student/berufstätig	11.5%	6.5%	6.6%
	Alten- oder Krankenhilfeschüler		93.5%	93.3%

Anmerkungen:

Stichprobenbeschreibung nach Ausschluss von Probanden aufgrund einer hohen Anzahl nicht bearbeiteter Items, geringer Sprachkenntnisse oder beeinträchtigende Sehbehinderung.

Abweichungen der aufsummierten Prozentpunkte zu 100% kommen durch fehlende Angaben bzw. Rundung zustande.

Versuchsablauf. Die Untersuchung 1 fand im Mai und Juni 2008 in Seminarräumen des Psychologischen Instituts und des Hörsaalgebäudes der Philipps-Universität Marburg in der Regel als Gruppentestung (1-13 Teilnehmer) statt. Die Testungen dauerten etwa 4 Stunden und wurden von zwei geschulten Testleiterinnen mit Unterstützung von zwei Psychologiestudentinnen

durchgeführt. Der Ablauf der Testungen wurde standardisiert durch ein Testleiter-Instruktionsheft mit genauen Angaben zu Ablauf, Instruktion und Verhalten bei möglichen Nachfragen. Die Videos des ViTEIP wurden via Beamer auf eine Leinwand projiziert und über Lautsprecher die begleitende Tonspur abgespielt. Die Beantwortung der Fragen erfolgte in einem ausgedruckten Testheft. Alle anderen Verfahren wurden als Papier-Bleistift-Versionen dargeboten.

Die insgesamt 16 Gruppentestungen (Gruppengröße 9-35 Personen) der Untersuchungen 2 wurden im Zeitraum Februar bis Mai 2009 von einer geschulten Testleiterin durchgeführt. Die Testungen wurden nach einem standardisierten Vorgehen in Klassenräumen der teilnehmenden Altenhilfe- und Krankenpflegeschulen durchgeführt und dauerten ca. 120 Minuten; daran anschließend wurde für die Teilnehmer eine kurze Schulungseinheit zum Thema Emotionale Intelligenz durchgeführt. Am Ende der Testung erhielt jeder Teilnehmer drei frankierte und adressierte Briefumschläge mit je einem EKF-Fremd Fragebogen, einem demografischen Fragebogen sowie dem Bogen „externe Kriterien“. Die Schüler wurden gebeten, diese Fragebögen von zwei Kollegen/Praxisanleitern und einem Klassenkameraden bearbeiten und an die Testleiterin zurücksenden zu lassen. Die Hälfte der Teilnehmer jeder Testgruppe bearbeitete den ViTEIP-V2a (Teilstichprobe 2a), die andere Hälfte den ViTEIP-V2b (Teilstichprobe 2b).

3.2 Testauswahl und –beschreibung

In *Tabelle 6* sind alle eingesetzten Verfahren, die zu erfassenden Konstrukte, die durchschnittliche Bearbeitungsdauer und die spezifische Untersuchung aufgeführt. Alle Verfahren sind Papier-Bleistift-Versionen und somit für die Durchführung von Gruppentestungen geeignet. Nachfolgend werden die Messinstrumente kurz erläutert, für eine ausführlichere Beschreibung sei jeweils auf die entsprechenden Manuale verwiesen. Für die Untersuchungen der Studie 2 standen pro Testung nur maximal drei Schulstunden zur Verfügung, so dass eine Auswahl der zu bearbeitenden Verfahren vorgenommen werden musste. Da im Zusammenhang mit Emotionaler Intelligenz zu Intelligenz und Persönlichkeitsmerkmalen schon diverse Studien vorliegen (vgl. Kapitel 1.4), wurden entsprechende Tests nur in Stichprobe 1 vorgegeben. Der Fokus der Studie 2 lag auf der Erhebung von Daten, die zur konvergenten Validierung (Emotionale Kompetenz, Empathie) und zur Einschätzung von Kriteriumsvalidität (Fremdeinschätzung Emotionale Kompetenz und Berufserfolg) geeignet sind. Die Untersuchung der Scorings erfolgt zur besseren Vergleichbarkeit und in der Erwartung der Verallgemeinerbarkeit der Ergebnisse anhand des neu entwickelten ViTEIP (siehe 1.5) und eines weiteren Tests zur Erfassung von Emotionaler Intelligenz, dem TEMINT (siehe 1.3.2).

Tabelle 6: *Eingesetzte Testverfahren in den Untersuchungen 1 und 2.*

Name des Tests	Zu erfassendes Konstrukt	Testart	Dauer ¹	U1	U2
ViTEIP	Emotionale Intelligenz	LT			
Version V1			60	✓	
textbasiert			20	✓	
Version V2a bzw. V2b			45		✓
TEMINT	Emotionale Intelligenz	LT	15	✓	✓
d2-R	Konzentrierte Aufmerksamkeit	LT	10	✓	✓
NEO-FFI	Persönlichkeit BIG 5	FB	15	✓	
IST-2000-R	Intelligenz	LT	70	✓	
Verbales Schlussfolgern					
Figurales Schlussfolgern					
IST-2000-R Verbale Merkfähigkeit	Merkfähigkeit	LT	10	✓	
WIT-2	Merkfähigkeit	LT	10		✓
EKF-Selbst	Emotionale Kompetenz	FB	15	✓	✓
Situationsbezogener Empathiefragebogen	Empathie	FB	20		✓
Demographischer Fragebogen	Demographische Daten	FB	5	✓	✓
EKF-Fremd²	Emotionale Kompetenz – Fremdeinschätzung	FB			✓
Externe Kriterien²	Berufserfolg – Fremdeinschätzung	FB			✓
Demographischer Fragebogen – Fremd²	Demographische Daten	FB			✓

Anmerkungen:

U1= Einsatz des Verfahrens in Untersuchung 1, überwiegend studentische Stichprobe; U2= Einsatz des Verfahrens in Untersuchung 2, überwiegend Altenhilfeschüler; LT= Leistungstest; FB= Fragebogen.

ViTEIP= Videobasierter Test zur Erfassung der Emotionalen Intelligenz im Pflegeberuf. TEMINT= Test zur Erfassung der Emotionalen Intelligenz (Schmidt-Atzert & Bühner, 2002). d2-R= Aufmerksamkeits-Konzentrations-Test d2, revidierte Form (Brickenkamp, Schmidt-Atzert, & Liepmann, 2010). NEO-FFI= NEO-Fünf-Faktoren Inventar nach Costa und McCrae (Borkenau & Ostendorf, 1993). IST-2000-R= Intelligenz-Struktur-Test 2000-R (Amthauer, Brocke, Liepmann, & Beauducel, 2001). WIT-2= Der Wilde-Intelligenztest (Kersting, Althoff, & Jäger, 2008). EKF= Emotionale-Kompetenz-Fragebogen (Rindermann, 2007). Situationsbezogener Empathiefragebogen= situationsbezogenen Fragebogen zur Erfassung von Empathie (Holz-Ebeling & Steinmetz, 1994).

Beschreibung der Verfahren siehe Text.

¹ Durchschnittliche Bearbeitungsdauer in Minuten.

² Diese Fragebögen wurden den Probanden ausgeteilt mit der Bitte, sie von einem Klassenkameraden und zwei Arbeitskollegen bearbeiten zu lassen.

Videobasierter Test zur Erfassung der Emotionalen Intelligenz im Pflegeberuf (ViTEIP). Die eingesetzten Versionen des ViTEIP wurden in Kapitel 1.5 beschrieben. Die Stichprobe 1 bearbeitete die Version V1, die Probanden der zweiten Untersuchung je zur Hälfte die Versionen V2a und V2b. Für die Vergleichbarkeit der Ergebnisse wurden für Stichprobe 1 nur die Items in die Berechnungen aufgenommen, die in der Version V2 inkludiert sind. Um den Einfluss des Präsentationsmodus zu prüfen, wurde eine textbasierte Forschungsversion entwickelt. Hier werden anstelle des Videos mit Sprecherin, alle Informationen auf einem Papierausdruck in Textform dargeboten. Diese Version besteht nur aus den Situationen 1, 3, 4, 6 und 9.

Fragen zur Akzeptanz des Verfahrens (Sozialen Validität). In Anlehnung an Kanning et al. (Kanning, Grewe, Hollenberg, & Hadouch, 2006) wurden den Probanden nach der Bearbeitung des ViTEIP

untenstehende Fragen zur Beantwortung auf einer vierstufigen Likertskala („ich stimme nicht zu“ bis „ich stimme zu“) vorgegeben. Mit diesen Fragen sollte die Akzeptanz (soziale Validität) des ViTEIP untersucht werden:

1. *Dieses Testformat (Situationsbeschreibung in Video- bzw. Textformat und Emotionseinschätzung) liefert nützliche Informationen für eine Personalauswahl im Pflegebereich.*
2. *Das Bearbeiten dieser Aufgaben hat mir gefallen.*
3. *Ich könnte dieses Testformat (Situationsbeschreibung in Video- bzw. Textformat und Emotionseinschätzung) in einem Auswahltest akzeptieren.*
4. *Ich halte einen Auswahltest mit diesem Testformat (Situationsbeschreibung in Video- bzw. Textformat und Emotionseinschätzung) für fair.*

Test zur Erfassung der Emotionalen Intelligenz (TEMINT). Die Beschreibung des TEMINT findet sich in Kapitel 1.3.2. Für den TEMINT liegen Ergebnisse zur konvergenten und divergenten Validität vor, z.B. von Schmidt-Atzert und Bühner (2002) oder von Blickle und Kollegen (Blickle G. , Momm, Liu, Witzki, & Steinmayr, 2011). Zur besseren Vergleichbarkeit der Ergebnisse wurden im TEMINT anstelle von Abweichungswerten analog des Scorings im ViTEIP Übereinstimmungswerte verwendet. Eine Beschreibung der eingesetzten Scorings erfolgt in Kapitel 3.3.1.

Emotionale-Kompetenz-Fragebogen (EKF). Der Emotionale-Kompetenz-Fragebogen (Rindermann, 2007) wurde in beiden Untersuchungen als Selbsteinschätzungs- und in Untersuchung 2 zusätzlich als Fremdbeurteilungsverfahren eingesetzt. Die Konzeptualisierung von Emotionaler Kompetenz erfolgt über vier Facetten: die Fähigkeit zum Erkennen eigener Gefühle (EE), die Fähigkeit zum Erkennen fremder Gefühle (EA), die Fähigkeit zur Regulation eigener Gefühle (RG) und die Fähigkeit zum Ausdruck von Gefühlen als emotionale Expressivität (EX). Den Probanden wurden 62 Aussagen („Auch in schwierigen Situationen kann ich mich gut steuern“ oder „Manchmal bin ich traurig, ohne zu wissen warum“) präsentiert, die auf einer fünfstufigen Skala („stimmt überhaupt nicht“ bis „stimmt vollständig“) bezüglich des Zutreffens auf die eigene (Selbstbeurteilung) bzw. die einzuschätzende Person (Fremdbeurteilung) beurteilt werden.

Aufmerksamkeits-Konzentrations-Test d2, revidierte Version (Test d2-R). Als Verfahren zur Erfassung der konzentrierten Aufmerksamkeit wurde eine standardisierte Weiterentwicklung (Brickenkamp, Schmidt-Atzert, & Liepmann, 2010) des Durchstreichtests d2 von Brickenkamp (2002) eingesetzt. Die revidierte Form des Tests enthält im Vergleich zur vorherigen Version eine anwenderfreundlichere Instruktion, eine größere Itemanzahl und eine leicht veränderte Auswertesystematik, die zur Erhöhung der Durchführungsobjektivität beiträgt. Die Probanden erhielten einen Arbeitsbogen, auf dem zwei Übungszeilen und 14 Zeilen mit je 57 Zeichen so schnell und so genau

wie möglich bearbeitet werden mussten. Aufgabe war es, eine Reizselektion vorzunehmen und aus den Distraktoren alle Zielreize (d mit zwei Strichen) zu identifizieren und durchzustreichen. Aus den bearbeiteten Zeichen (Arbeitstempo) und aus der Genauigkeit der Bearbeitung (Abzug von Auslassungs- und Verwechslungsfehlern) wurde der Kennwert Konzentrationsleistung berechnet.

NEO-Fünf-Faktoren-Inventar (NEO-FFI). Die Stichprobe 1 bearbeitete ein Selbstbeschreibungsverfahren zur Erfassung von Persönlichkeitsmerkmalen. Das NEO-FFI nach Costa und McCrae (Borkenau & Ostendorf, 1993) ist ein multidimensionales Persönlichkeitsinventar und erfasst die Skalen Neurotizismus, Extraversion, Offenheit für Erfahrung, Verträglichkeit und Gewissenhaftigkeit. Die insgesamt 60 Items wurden als Aussagen präsentiert, die von den Probanden nach dem Zustimmungsgrad (fünfstufig von „starke Ablehnung“ bis „starke Zustimmung“) beurteilt werden.

Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R). Als Intelligenzscreening wurden in Studie 1 die Untertests *Verbale Intelligenz* und *Figurale Intelligenz* des Intelligenz-Struktur-Test 2000-R (I-S-T 2000 R) von Amthauer, Brocke, Liepmann und Beauducel (2001) eingesetzt. Für beide Facetten kamen je drei Aufgabentypen à 20 Items zum Einsatz: Satzergänzung, Analogien und Gemeinsamkeiten für die Erfassung der verbalen Intelligenz sowie Figurenauswahl, Würfelaufgaben und Matrizen für die Erfassung der figuralen Intelligenz. Nach einer Erläuterung der Aufgaben durch den Testleiter bearbeiteten die Probanden die Aufgaben selbstständig in einer vorgegebenen Zeit. Die Beantwortung erfolgte im Multiple-Choice-Format.

Verfahren zur Erfassung von Merkfähigkeit (I-S-T 2000 R Untertest Merkfähigkeit bzw. WIT-2 Untertest Merkfähigkeit). In Studie 1 wurde die Gedächtnisleistung mit dem Untertest Merkfähigkeit (verbal) des Intelligenz-Struktur-Tests 2000 R (I-S-T 2000 R) von Amthauer, Brocke, Liepmann und Beauducel (2001) erfasst. In einer 1-minütigen Lernphase mussten vorgegebene Wörter und entsprechende Zuordnung zu Oberbegriffen eingeprägt werden. In der Abrufphase wurden nur die Anfangsbuchstaben der gelernten Wörter präsentiert; die Zuordnung zu den Oberbegriffen war Aufgabe der Probanden. Aufgrund von Deckeneffekten in Stichprobe 1 wurde in Studie 2 ein anderes Verfahren zur Erfassung von Merkfähigkeit eingesetzt, der Subtest Merkfähigkeit des Wilde-Intelligenz-Test 2 (WIT-2) von Kersting, Jäger und Althoff (2008). Hier lernten die Probanden in der Einprägephase im Rahmen einer Geschichte verschiedene Informationen, die per Multiple-Choice-Fragen in der Reproduktionsphase abgefragt wurden.

Empathiefragebogen. Als verwandtes Konstrukt zu Emotionaler Intelligenz wurde Empathie mittels des situationsbezogenen Fragebogens von Holz-Ebeling und Steinmetz (1994) erfasst. Empathie wird hierbei verstanden als das „Erfassen (Begreifen und/oder Nacherleben) fremder Bewusstseinsinhalte (Gedanken, Gefühle und/oder Bedürfnisse) im Kontext bestimmter Lebensumstände...“ (Holz-

Ebeling & Steinmetz, 1994, S. 159). Den Probanden wurden kurze Situations-beschreibungen (z.B. „Als Sie mit dem Bus in die Stadt fahren, beobachten Sie, dass eine überarbeitet und müde aussehende Frau einem alten Mann aus Versehen auf den Fuß tritt. Er schnauzt sie so grob an, dass sie ganz erschrocken guckt.“) präsentiert mit der Frage, wie stark sie am Ende der Schilderung den Eindruck haben zu wissen, was in der anderen Person gerade vor sich geht. Auf einer siebenstufigen Skala („überhaupt nicht“ bis „sehr stark“) gaben die Probanden ihre Einschätzung wieder. Für die Vollversion des Tests geben die Autoren eine interne Konsistenz von $\alpha=0.87$ an. In der vorliegenden Arbeit wurden aus zeitökonomischen Gründen nur die 16 Items zur Erfassung kognitiver Empathie¹⁷ zur Bearbeitung vorgegeben.

Demografischer Fragebogen. In einem kurzen Fragebogen wurden verschiedene Angaben zur Person (Alter, Muttersprache bzw. Deutschkenntnisse, Beruf oder Fachrichtung des Studiengangs, Semesterzahl bzw. Ausbildungsjahr) und zu Erfahrungen im Pflegebereich sowie im Umgang mit älteren Menschen erhoben.

Fragebogen „Externe Kriterien“. Der Fragebogen „Externe Kriterien“ diente in Stichprobe 2 dazu, für die Teilnehmer eine Einschätzung durch Kollegen, Praxisanleiter und Klassenkameraden in Bezug auf die Eignung für den Pflegeberuf zu erheben. Hierzu wurde die Eignung für verschiedene Aufgabenbereiche (z.B. Umgang mit Bewohnern, Büro und Verwaltungsarbeiten, Grundpflege der Bewohner, Beratung der Angehörigen, Körperliche Eignung, Umgang mit Kollegen) sowie eine Einschätzung der allgemeinen beruflichen Eignung abgefragt. Die Beantwortung erfolgt mittels einer sechsstufigen Antwortskala (1= sehr gut bis 6=ungenügend).

3.3 Vorgehen bei der Datenauswertung

Die Dateneingabe erfolgte im Programm „Statistical Package for Social Sciences“ (SPSS) für Windows, Version 15. Die korrekte Eingabe und Auswertung der Daten wurde mehrfach durch das Testungsteam und mithilfe von Plausibilitätschecks überprüft. Die statistischen Analysen wurden mit SPSS 19 bzw. SPSS 20 sowie Excel 2010 durchgeführt.

3.3.1 Auswertung der Emotionale-Intelligenz-Leistungstests

Die Rohwerte der Probandenangaben im ViTEIP-V2 und TEMINT wurden zunächst manuell in SPSS übertragen und kodiert. Im ViTEIP ergibt die angekreuzte Antwortoption „Emotion nicht

¹⁷ Die Autoren unterscheiden zwischen „kognitiver“ und „affektiver“ Empathie und beziehen sich damit darauf, wie eine Person die Emotionen und Gedanken einer Originalperson erfasst. „Kognitive Empathie“ geht mit dem Erkennen und Verstehen von Emotionen anderer einher; „affektive Empathie“ mit dem Erleben und Nachfühlen der Emotionen anderer (Holz-Ebeling & Steinmetz, 1994).

vorhanden“ den Zahlenwert 0, die Antwortoption „Emotion schwach vorhanden“ den Zahlenwert 1 usw. Entsprechend ergibt sich beim TEMINT für die Antwortoption „Emotion nicht oder sehr schwach vorhanden“ der Zahlenwert 0, bei „Emotion schwach bis mittel vorhanden“ der Wert 1 usw. Die Auswertungen erfolgten auf Itemebene bezogen auf jede einzelne Emotion.

In der Literatur werden diverse Möglichkeiten vorgestellt, wie die Kennwerte generiert werden können. Eine differenzierte und umfangreiche Untersuchung verschiedener Konsensansätze führte MacCann (MacCann, Roberts, Matthews, & Zeidner, 2004) durch; alternative Ansätze im Rahmen des Targetscorings stellen beispielsweise Berechnungen der Euklidischen Distanz oder gewichtete Abweichungswerte dar, wie sie in der Arbeitsgruppe von Süß zu sozialer Intelligenz untersucht werden (Baumgarten, persönliche Kommunikation am 14.12.2011). Da es in der vorliegenden Arbeit darum geht, die beiden Scoringansätze Target und Konsens in ihren grundsätzlichen Eigenschaften miteinander zu vergleichen, wurden innerhalb der Ansätze jeweils zwei verschiedene Herangehensweisen untersucht. Mit Target-Übereinstimmung und Konsens-Anteil wurden Scorings gewählt, bei der auch annähernd richtige Antworten honoriert werden; bei Target-Treffer und Konsens-Modal handelt es sich um „strengere“ Auswertungen, bei denen nur eine Antwortoption als richtig gewertet wird.

Target-Übereinstimmung (T-Ü). Zwischen den Angaben der Originalperson und den Angaben der Probanden wurde auf Itemebene die Übereinstimmung berechnet (siehe Abbildung 9). Je weiter die Antwort des Probanden von der Angabe der Originalperson entfernt liegt, desto geringer ist der Übereinstimmungswert. Antwortet ein Proband in Übereinstimmung mit der Antwort der Originalperson wurde dies mit dem Wert „3“ kodiert (in Abbildung 9, Emotionen „Ärger“ und „Freude“). Antwortet der Proband um eine Einheit abweichend nach rechts oder links, ergibt dies einen Übereinstimmungswert von „2“ (in Abbildung 9 Emotion „Angst/Furcht“ und „Schuld“).

Target-Treffer (T-T). Nur eine exakte Übereinstimmung zwischen Angabe der Originalperson und dem Urteil des Probanden wird als richtige Antwort mit dem Wert „1“ bewertet. In Abbildung 9 trifft dies auf die Emotionen „Ärger“ und „Freude“ zu. Im Vergleich zu Target-Übereinstimmung werden bei Target-Treffer jedoch nur einen Teil der in den Daten vorhandenen Informationen genutzt, was zu einer weniger differenzierten Auswertung führt.

Konsens-Anteil (K-A). Die Antworten der Probanden wurden entsprechend des Anteils an Personen gewertet, die dieselbe Antwort gaben. Bsp.: 20% der Probanden gaben Antwort A, 45% die Antwort B, 35% die Antwort C. Probanden mit der Antwort A erhalten einen Testwert von 0.20, diejenigen mit der Antwort B einen Wert von 0.45, mit Antwort C den Wert 0.35. Da sich der ViTEIP-V2 noch im

Entwicklungsstadium befindet, wurden die Anteilswerte an der jeweiligen Stichprobe selbst berechnet und nicht an einer zusätzlichen Normstichprobe.

Konsens-Modal (K-M). Nur die Antwortkategorie, die von den meisten Probanden gewählt wurde, wird als richtige Antwort gewertet und mit einem Wert von 1 verrechnet, alle anderen Antwortkategorien ergeben einen Wert von 0.

	nicht vorhanden	schwach vorhanden	mittel vorhanden	stark vorhanden	Target- Übereinstimmung	Target-Treffer
Abneigung	<input checked="" type="radio"/>	<input type="radio"/> ✓	<input type="radio"/>	<input type="radio"/>	2	0
Ärger	<input type="radio"/>	<input checked="" type="radio"/> ✓	<input type="radio"/>	<input type="radio"/>	3	1
Angst/Furcht	<input type="radio"/>	<input type="radio"/> ✓	<input checked="" type="radio"/>	<input type="radio"/>	2	0
Unruhe	<input type="radio"/>	<input type="radio"/> ✓	<input type="radio"/>	<input checked="" type="radio"/>	1	0
Traurigkeit	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/> ✓	0	0
Enttäuschung	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/> ✓	1	0
Schuld (-gefühl)	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/> ✓	2	0
Freude	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/> ✓	3	1

Anmerkungen:

☒ = Urteil des Probanden; ✓ = Angabe der Originalperson.

Links im Bild: Darstellung des Antwortbogens.

Rechts im Bild: Darstellung der berechneten Testkennwerte bei Anwendung des Scorings Target-Übereinstimmung und Target-Treffer.

Abbildung 9: Darstellung der Auswertung mittels Target-Übereinstimmung und Target-Treffer an einem fiktiven Beispiel

Fehlende Werte. Für den Umgang mit fehlenden Werten gibt es verschiedene Möglichkeiten. Ein mögliches Vorgehen ist es, unbearbeitete Items mit dem minimal möglichen Übereinstimmungswert zu ersetzen. Das bedeutet, dass als Antwort des Probanden fiktiv die Antwortoption gewählt wird, die am weitesten entfernt liegt von der Angabe der Originalperson. Für das in *Abbildung 9* dargestellte Beispiel würde ein fehlender Wert bei der Emotion „Abneigung“ mit der Antwortoption „stark vorhanden“ und bei der Emotion „Enttäuschung“ mit der Antwortoption „nicht vorhanden“ ersetzt. Dieses Vorgehen führt dazu, dass das Auslassen von Items für die Probanden keinen Vorteil in den Gesamtergebnissen ergibt; die Probanden sollten jedoch im Vorfeld der Testbearbeitung über die Vorgehensweise bezüglich nicht bearbeiteter Items informiert werden. Dieser Ansatz ist aus Sicht der Autorin das Vorgehen der Wahl, wenn der ViTEIP außerhalb von Forschungsarbeiten in der Praxis zum Einsatz kommt.

Damit der Vergleich zwischen den Antwortformaten „Ankreuzen“ und „Zuordnen“ (siehe Kapitel 1.5.3 und 2.2) möglich ist, wurde in den vorliegenden Studien eine andere Alternative gewählt:

fehlende Werte wurden mit der Antwort „Emotion nicht vorhanden“ ersetzt. Dies hat den Vorteil, dass ein vergleichbares Vorgehen bei beiden Antwortformaten gewährleistet ist. Anders als bei „Ankreuzen“ gibt es bei „Zuordnen“ keine Missings, da diese nicht von den Items separiert werden können, die die Probanden im Bearbeitungsfeld stehen lassen, um damit zum Ausdruck zu bringen „Emotion nicht vorhanden“.

3.3.1.1 Standardisierung der Daten für den Vergleich der Auswertungsstrategien

Um für den Vergleich der Scorings (Hypothesen 1a und b; Stichproben 1 und 2a¹⁸) bei beiden Tests und allen Auswertungsprozeduren denselben Maßstab zu haben, wurden die Messwerte (X_i) linear transformiert (Bühner, 2006). Die Transformation erfolgte auf Itemebene; also auf der Ebene der einzelnen Emotionen (Angst, Trauer, etc.). Die Itemebene ist von „Situationen“ abzugrenzen, die aus mehreren Items bestehen und inhaltlich zusammengehörige Cluster von Emotionen bilden. Nachvollzogen werden kann diese Unterscheidung in *Abbildung 9*, wo die einzelnen Zeilen mit den Emotionen den Items entsprechen und die Kombination aller aufgeführten Emotionen die Situation darstellt. Für die lineare Transformation wurden die Werte verwendet, die sich nach Anwendung der verschiedenen Auswertungsstrategien ergeben (z.B. bei Scoring Target-Übereinstimmung: die Übereinstimmungswerte auf Itemebene). Über eine z-Transformation wurde die Standardabweichung als Maßeinheit ($SD=1$) und der Mittelwert als neuen Nullpunkt ($M=0$) definiert, um so die Ausprägung in einem einzelnen Item über die relative Lage zu den Ausprägungen der Probanden der Stichprobe in Bezug auf dieses jeweilige Item miteinander zu vergleichen. Anschließend wurde für jede eingesetzte Auswertungsprozedur ein separater Kennwert aus der Summe der z-transformierten Items gebildet. 12 Items des TEMINT wurden bei der Berechnung des Gesamtwertes nicht berücksichtigt, da diese Items in mindestens einer Stichprobe von allen Probanden einheitlich (d.h. 100% Übereinstimmung in den Antworten) beantwortet wurden und somit keine Varianz aufwiesen (siehe *Tabellen A-3 und A-4 in Anhang C* zur prozentualen Verteilung der Antworten auf die Antwortoptionen). In den Gesamtwert des ViTEIP fließen 99 Items ein, in den Gesamtwert des TEMINT 73 Items. Zur Reliabilitätsschätzung wurden Cronbachs α und die Testhalbierungs-Methode (Aufteilung nach Odd-Even¹⁹) eingesetzt.

¹⁸ Stichprobe 2b wird für diese Fragestellung nicht herangezogen, da diese Teilnehmergruppe den ViTEIP-V2b mit einem veränderten Antwortformat bearbeitet und somit andere Testbedingungen vorliegen.

¹⁹ Aufteilung der Items mit gerader bzw. ungerader Reihungsnummer in zwei Testhälften.

3.3.2 Auswertungen der weiteren eingesetzten Verfahren

Die Auswertungen der eingesetzten Verfahren wurden entsprechend der Angaben in den Testmanualen durchgeführt.

Die Auswertung des *Emotionale-Kompetenz-Fragebogens* (Rindermann, 2007) erforderte im ersten Schritt eine Kodierung der Antworten (von „stimmt überhaupt nicht“=1 bis „stimmt vollständig“=5). Die kodierten Probandenantworten wurden in SPSS übertragen und mithilfe einer Syntax zusammengeführt. Invers formulierte Items wurden über die Syntax umgepolt. Die den vier Skalen „Erkennen eigener Emotionen“, „Erkennen von Emotionen bei anderen“, „Regulation und Kontrolle eigener Emotionen“ und „Emotionale Expressivität“ zugehörigen Items wurden jeweils aufsummiert und durch die Anzahl der bearbeiteten Items dieser Dimension geteilt. Die Skalenmittelwerte wurden z-transformiert und zum Gesamtwert „Allgemeine Emotionale Kompetenz“ aufsummiert. Die Auswertung des EKF-Fremd erfolgte analog des EKF-Selbst.

Die Untertests des *IST-2000-R* (Amthauer, Brocke, Liepmann, & Beauducel, 2001) wurden mithilfe von Schablonen ausgewertet. Es wurden Rohpunktsummen für die Untertests zur verbalen und figuralen Intelligenz sowie zur Merkfähigkeit gebildet.

Kennwert zum Merkfähigkeitstest *WIT-2* (Kersting, Althoff, & Jäger, 2008) ist die Anzahl der richtig erinnerten Antworten. Die Auszählung erfolgte manuell; Gesamtwertwerte können zwischen 0 und 21 liegen.

Für den *d2-R* (Brickenkamp, Schmidt-Atzert, & Liepmann, 2010) wurden mittels Auswertungsschablone die Gesamtzahl der bearbeiteten Zeichen ausgezählt und die Auslassung- und Verwechslungsfehler identifiziert. Der in dieser Arbeit verwendete Kennwert Konzentrationsleistung berechnete sich aus der Differenz der Auslassung- und Verwechslungsfehler von der Gesamtzahl der bearbeiteten Zeichen.

Die Auswertung des *NEO-FFI* (Borkenau & Ostendorf, 1993) fand anhand von Schablonen statt. Für jeden Proband wurden die Rohwerte der fünf Skalen Neurotizismus, Extraversion, Offenheit für Erfahrungen, Verträglichkeit, Gewissenhaftigkeit ermittelt.

Die Auswertung des *Empathiefragebogens* (Holz-Ebeling & Steinmetz, 1994) erfolgte durch eine Summierung der gegebenen Antworten. Der Kennwert zu Empathie liegt zwischen 16 und 112.

Die Antworten des *Fragebogens zu externen Kriterien* wurden mit einer entsprechenden numerischen Kodierung in SPSS übertragen (z.B. Vorgesetzter=1, Kollege=2, Ausbilder/Praxisanleiter=3, anderer beruflicher Kontext=4). Direkt übernommen werden konnten die Angaben zur Eignung als Pflegekraft und der Erfüllung der Berufsanforderungen.

4 Ergebnisse

Nachfolgend werden die Ergebnisse der zentralen Berechnungen und Analysen vorgestellt. Aus Gründen der Übersichtlichkeit werden an einigen Stellen nur die Ergebnisse einer Stichprobe dargestellt; die Ergebnisse der weiteren Stichproben sind im Anhang aufgeführt. Von einer Zusammenfassung der Daten über alle Stichproben hinweg wurde abgesehen, da sich die Stichproben 1 und 2 in der Probandenzusammensetzung (Studenten vs. Altenhilfeschüler) sowie den ViTEIP-Antwortformaten (Ankreuzen vs. Zuordnen) deutlich voneinander unterscheiden.

Eine Tabelle der deskriptiven Kennwerte der Verfahren, die zusätzlich zu den Emotionale-Intelligenz-Tests eingesetzt wurden, findet sich in *Anhang D, Tabelle A-5*. Im Vergleich zu der Stichprobe der Pflegeschüler schnitt die studentischen Stichprobe 1 im Test d2-R (Effektstärke $d \geq 0.71$) deutlich besser ab. Die Unterschiede in einigen EKF-Facetten liegen mit Effektstärken $d \leq 0.03$ im sehr geringen Bereich.

In der zweiten Untersuchung erhielt jeder Teilnehmer drei adressierte und frankierte Umschläge mit Fragebögen zur Fremdeinschätzung mit der Bitte um Weitergabe an Kollegen und Klassenkameraden. Insgesamt wären bei einem vollständigen Rücklauf etwa 900 Fremdeinschätzungen möglich; trotz mehrfacher Erinnerung per Email wurden nur 40 Bögen vollständig ausgefüllt zurückgesendet. Die Quote der Rückantworten lag damit deutlich unter der aufgrund von Erfahrungen erwarteten Rücklaufquote von 30%. Aufgrund dieses geringen Rücklaufs konnten die ursprünglich vorgesehenen Berechnungen zur Kriteriumsvalidität (Außenkriterien und Fremdeinschätzungen) nicht durchgeführt werden.

4.1 Vergleich der Auswertungsstrategien

Die Emotionale-Intelligenz-Tests wurden daraufhin untersucht, in welcher Form sich die Anwendung verschiedener Scorings auf die Ergebnisse auswirkte. Die Berechnungen wurden an den Ergebnissen der studentischen Stichprobe 1 und der Teilstichprobe 2a (Altenhilfeschüler) durchgeführt, die den ViTEIP-V2 mit demselben Antwortformat (Ankreuzen) bearbeiteten.

4.1.1 Deskriptive Statistiken und Exploration der faktoriellen Struktur

Unter targetbasierten Auswertungen werden die Scorings Target-Übereinstimmung und Target-Treffer verstanden, bei denen sich die Testwerte daraus generieren, dass die Angabe des Probanden mit der Angabe der Originalperson – also dem Target – verglichen und ein Übereinstimmungs- bzw. Trefferwert berechnet wird. Bei den konsensbasierten Scorings Konsens-Anteil

und Konsens-Modal wird der Gesamtwert aus der Übereinstimmung mit der Mehrheitsmeinung der Stichprobe berechnet. Eine genaue Beschreibung der Scorings findet sich in Kapitel 3.3.1. Damit bei allen vier Auswertungsverfahren dieselbe Skalierung vorliegt und eine direkte Vergleichbarkeit der Ergebnisse möglich war, wurde über eine z-Transformation die Maßeinheit der Standardabweichung als $SD=1$ und der Mittelwert $M=0$ definiert. Stichprobenmittelwerte sind in *Tabelle 7* nicht angegeben, da diese aufgrund der z-Transformation qua definitionem bei $M=0.00$ liegen.

Tabelle 7: Deskriptive Statistiken ViTEIP-V2a und TEMINT bei verschiedenen Scorings

Test	Stichprobe	Scoring	s^2	r_{tt}	α	Kolmogorov-Smirnov
ViTEIP-V2a	1	Target-Übereinstimmung	306.29	0.76	0.70	0.07 ($p=0.20$)*
		Target-Treffer	275.04	0.66	0.65	0.05 ($p=0.20$)*
		Konsens-Anteil	249.44	0.72	0.61	0.08 ($p=0.08$)
		Konsens-Modal	218.53	0.63	0.55	0.06 ($p=0.20$)*
	2a	Target-Übereinstimmung	330.56	0.77	0.71	0.06 ($p=0.20$)*
		Target-Treffer	181.27	0.53	0.46	0.05 ($p=0.20$)*
		Konsens-Anteil	178.81	0.44	0.45	0.05 ($p=0.20$)*
		Konsens-Modal	149.62	0.32	0.34	0.06 ($p=0.20$)*
TEMINT	1	Target-Übereinstimmung	314.70	0.80	0.79	0.11 ($p=0.00$)
		Target-Treffer	274.00	0.80	0.75	0.09 ($p=0.02$)
		Konsens-Anteil	214.46	0.74	0.67	0.10 ($p=0.01$)
		Konsens-Modal	188.40	0.73	0.63	0.10 ($p=0.01$)
	2a	Target-Übereinstimmung	247.64	0.76	0.72	0.08 ($p=0.02$)
		Target-Treffer	242.16	0.75	0.71	0.08 ($p=0.04$)
		Konsens-Anteil	171.60	0.60	0.60	0.10 ($p=0.00$)
		Konsens-Modal	183.03	0.68	0.63	0.09 ($p=0.01$)

Anmerkungen:

Scoring: Für eine detaillierte Beschreibung der Scorings siehe Kapitel 3.3.1.

Verwendung der auf Itemebene standardisierten Werte, Mittelwerte sind in der Tabelle nicht angegeben, da diese aufgrund der z-Transformation auf Stichprobenebene qua definitionem bei $M=0.00$ liegen, siehe Kapitel 3.3.1.1.

s^2 = Varianz. r_{tt} = Split-Half-Reliabilität, Aufteilung des Tests nach der Odd-Even-Methode, Schätzung der Reliabilität des Gesamttests nach Guttman. α = interne Konsistenz Cronbachs α . Kolmogorov-Smirnov= Kolmogorov-Smirnov-Test auf Normalverteilung mit Signifikanzkorrektur nach Lilliefors.

* Dies ist eine Untergrenze der echten Signifikanz.

Es ist zu erkennen, dass sich bei beiden Emotionale-Intelligenz-Tests und in beiden Stichproben die Varianz bei verschiedenen Scorings unterscheidet; bei allen targetbasierten Auswertungen fiel diese deutlich größer aus als bei konsensbasierten Scorings. Als Beispiel für diese Varianzunterschiede soll hier exemplarisch die Stichprobe 1 (ViTEIP-V2a) mit einer Varianz $s^2=306.29$ bei Target-Übereinstimmung und einer deutlich geringeren Varianz von $s^2=218.53$ bei Konsens-Modal genannt werden. Sowohl die Reliabilitätsschätzung mittels Cronbachs α als auch mittels Testhalbierungsmethode²⁰ ergaben für das Scoring Target-Übereinstimmung die höchsten Werte: Diese lagen beim ViTEIP-V2a über $\alpha \geq 0.70$ und beim TEMINT über $r_{tt} \geq 0.76$. Geringer fielen die Reliabilitäten für

²⁰ Reliabilitätsschätzung nach Guttman, da sich die Varianzen in beiden Testhälften unterscheiden.

konsensbasierte Auswertungen aus; hier lagen die Werte zwischen $r_{tt}=0.32$ (ViTEIP-V2a Konsens-Modal) und maximal $r_{tt}=0.74$ (TEMINT Konsens-Modal). Ins Auge stechen die ViTEIP-Reliabilitäten in Stichprobe 2a, die außer bei Target-Übereinstimmung vergleichsweise gering sind. Warum die Reliabilität (sowie auch die Varianz) gerade in dieser Teilstichprobe für den ViTEIP so gering ausfällt, lässt sich aus den vorliegenden Daten nicht erklären: Es liegen keine Eingabefehler vor, die Stichprobe weicht in der Zusammensetzung nicht von den anderen Stichproben ab und die Leistungen in anderen Verfahren sind vergleichbar. Explorative Untersuchungen, ob andere Variablen (z.B. Alter, Anzahl der Missings in den Tests, Ausbildungsjahrgang,...) im Sinne von Störvariablen oder Subgruppenbildungen die Ergebnisse beeinflussen, blieben ohne Ergebnis.

Insgesamt führte die Anwendung eines anderen Scorings zu Unterschieden in der Höhe der Varianz und der Reliabilität. Bei ViTEIP-V2a und TEMINT zeigten sich in beiden Stichproben bei dem Scoring Target-Übereinstimmung die größte Varianz und die höchste Reliabilität.

Exploration der faktoriellen Struktur des ViTEIP-V2a und TEMINT. Als nächstes wurde explorativ überprüft, welche Auswirkungen die Anwendung eines anderen Scorings auf die faktorielle Zuordnung der Items hat. Dazu sollte mit Faktorenanalysen berechnet werden, wie viele Faktoren den Items zugrunde liegen und wie sich die Items auf diese Faktoren verteilen. Da es sich mit ViTEIP-V2a und TEMINT jeweils um Tests handelt, die ein einheitliches Konstrukt (bzw. einen Teilbereich eines Konstrukts) erfassen sollen, war zu erwarten, dass es einen starken ersten Faktor gibt, der einen Großteil der Varianz erklärt und sich Items bei unterschiedlichen Scorings auf denselben Faktoren wiederfinden. Als faktorenanalytische Methode wurden Hauptkomponentenanalysen mit Promax-Rotationen durchgeführt. Dieses Rotationsprinzip stellt ein Vorgehen dar, bei dem zunächst mit einer orthogonalen Rotation gestartet und anschließend zu einer obliquen (korrelierten) Lösung übergegangen wird. Primärladungen werden dabei maximiert und Sekundärladungen minimiert.

Da der Kaiser-Meyer-Olkin-Koeffizient (KMO) sowohl bei ViTEIP-V2a als auch TEMINT deutlich unter der empfohlenen Grenze von 0.60 lag (Bühner, 2006, S. 210), war auf Itemebene eine faktorenanalytische Untersuchung nicht möglich. Daher wurden die Items nach ihrer inhaltlichen Zugehörigkeit zu den Situationen²¹ zusammengefasst. Für den ViTEIP-V2 konnten nun aufgrund eines KMO knapp über der empfohlenen Grenze von 0.60 für die Auswertungen mit Target-Übereinstimmung (Stichprobe 1 und 2a), Target-Treffer (nur Stichprobe 1) und Konsens-Anteil (nur Stichprobe 1) Hauptkomponentenanalysen gerechnet werden. Für den TEMINT konnten aufgrund der Dateneignung in Stichprobe 1 und 2a Hauptkomponentenanalysen für Target-Übereinstimmung

²¹ Situationen bestehen aus mehreren Items und bilden inhaltlich zusammengehörige Cluster von Emotionen, so wie sie von den Originalpersonen geschildert wurden.

und Target-Treffer berechnet werden; in Stichprobe 1 zudem für Konsens-Anteil. Für die Daten mit einem Scoring Konsens-Modal lag für beide Tests in beiden Stichproben keine Eignung für faktorenanalytische Untersuchungen vor. Die Struktur- und Komponentenkorrelations-Matrizen von Stichprobe 1 und 2a sind in *Anhang E, Tabellen A-6 bis A-23* abgebildet.

Es zeigte sich in allen durchgeführten Hauptkomponentenanalysen eine erste Komponente, die zwischen 17.69% und 26.10% der Varianz aufklärt, gefolgt von drei bis vier weiteren Komponenten mit geringeren Anteilswerten (zwischen 8.35% und 14.43%) der Varianzaufklärung. Dass eine erste Komponente mehr Varianz aufklärt als die weiteren, entspricht der Erwartung; aufgrund der vergleichsweise geringen Aufklärungshöhe und der Existenz weiterer, vergleichsweise starker Komponenten ist dennoch anzuzweifeln, ob mit ViTEIP-V2 und TEMINT homogene Konstrukte erfasst werden. Jenseits der Varianzaufklärung fiel bei beiden Tests und über die Stichproben hinweg auf, dass es deutliche Unterschiede darin gab, welche Situationen jeweils auf den Komponenten luden sowie welche Situationen die höchste Ladung auf der ersten Komponente aufwiesen. Die Verwendung unterschiedlicher Scorings führte somit dazu, dass sich die Zuordnung der Situationen auf die Komponenten veränderte.

4.1.2 Korrelationsanalysen

Die Korrelationen des ViTEIP-V2a und TEMINT untereinander sowie zu den in Stichprobe 1 und 2a bearbeiteten Verfahren wurde betrachtet.

Korrelative Zusammenhänge der Scorings innerhalb der Tests. Aus *Tabelle 8* lässt sich in den grau hinterlegten Zellen ablesen, dass die Gesamtergebnisse bei den verschiedenen Auswertungsstrategien innerhalb desselben Tests signifikant positiv miteinander korrelierten (Korrelationen $r \geq 0.48^{**}$). Eine Ausnahme bilden die Korrelationen zwischen den verschiedenen Scorings des ViTEIP-V2a in Stichprobe 2a; hier fielen die Zusammenhänge zwischen den konsens- und den targetbasierten Ansätzen deutlich geringer aus und liegen unter $r \leq 0.35^{**}$. Die engsten Zusammenhänge ($r \geq 0.70^{**}$) fanden sich für ViTEIP-V2a und TEMINT jeweils innerhalb derjenigen Auswertungsansätze, die auf dasselbe Grundprinzip (Target bzw. Konsens) zurückgehen. Vor allem beim TEMINT lag diese Korrelation mit Werten über $r \geq 0.95^{**}$ in einem sehr hohen Bereich.

Es erfolgte eine Überprüfung des ViTEIP-V2 und TEMINT im Sinne der konvergenten und diskriminanten Validität (*Tabelle 8 bzw. 9*). Anschließend wurde betrachtet, inwiefern die Anwendung eines anderen Scorings zu Veränderungen in der Korrelationshöhe führte (*Tabelle 10*).

Korrelation zwischen ViTEIP-V2 und TEMINT. Im Rahmen der Analyse der konvergenten Validität wurden die Korrelationen der Emotionale-Intelligenz-Tests untereinander betrachtet. Es zeigte sich

ein unerwartetes Ergebnis (siehe weiße Zellen in *Tabelle 8*), welches unter 4.2 differenzierter untersucht wird. Auf Gesamttestebene war die Korrelation zwischen ViTEIP-V2 und TEMINT bei targetbasierter Auswertung signifikant-negativ bzw. nicht signifikant (zwischen $r=-0.33^{**}$ und $r=-0.03$); bei konsensbasiertem Scoring hingegen fiel die Korrelation positiv aus (zwischen $r=0.24^{**}$ und $r=0.33^{**}$). Wird ein Test mit Konsens, der andere mit Target ausgewertet, waren die Korrelationen überwiegend nicht signifikant zwischen $r\geq-0.13$ und $r\leq 0.19^*$.

Tabelle 8: Korrelation von ViTEIP-V2 und TEMINT bei unterschiedlichen Auswertungsmethoden

Stich- probe	Test	Scoring	ViTEIP				TEMINT			
			Target- Übereinst.	Target- Treffer	Konsens- Anteil	Konsens- Modal	Target- Übereinst.	Target- Treffer	Konsens- Anteil	Konsens- Modal
1	ViTEIP	Target-Übereinst.								
		Target-Treffer	0.86**							
		Konsens-Anteil	0.63**	0.48**						
		Konsens-Modal	0.51**	0.52**	0.89**					
	TEMINT	Target-Übereinst.	-0.15	-0.07	-0.13	-0.09				
		Target-Treffer	-0.14	-0.03	-0.10	-0.05	0.97**			
		Konsens-Anteil	0.16*	0.18*	0.33**	0.32**	0.70**	0.66**		
		Konsens-Modal	0.17*	0.19*	0.32**	0.31**	0.71**	0.70**	0.97**	
2a	ViTEIP	Target-Übereinst.								
		Target-Treffer	0.70**							
		Konsens-Anteil	0.35**	0.17*						
		Konsens-Modal	0.09	0.22**	0.82**					
	TEMINT	Target-Übereinst.	-0.23**	-0.28**	0.08	0.07				
		Target-Treffer	-0.33**	-0.24**	0.05	0.11	0.93**			
		Konsens-Anteil	-0.01	-0.06	0.28**	0.25**	0.70**	0.67**		
		Konsens-Modal	-0.09	-0.06	0.24**	0.24**	0.70**	0.73**	0.95**	

Anmerkungen:

Korrelation nach Pearson. Einseitige Testung, * $p<0.05$. ** $p<0.01$.

Scoring: Für eine detaillierte Beschreibung der Scorings siehe Kapitel 3.3.1.

Für die Berechnungen wurden die standardisierten Werte verwendet, siehe Kapitel 3.3.1.1.

Aus Gründen der Übersichtlichkeit wurden die Zellen, in denen Korrelationen innerhalb eines Tests aufgeführt sind, grau hinterlegt.

Konvergente Validität. Die Korrelation zur verbalen Intelligenzfacette (*Tabelle 9*) fiel bei allen Scorings für den ViTEIP-V2a mit $r\geq 0.22^{**}$ signifikant und für den TEMINT nicht signifikant aus. In Stichprobe 1 zeigten sich für den ViTEIP-V2a einige positive signifikante Korrelationen zu Emotionaler Kompetenz (EKF, Gesamttest- oder/und Skalenebene) mit einer maximalen Korrelation von $r\leq 0.22^{**}$. In Stichprobe 2a fanden sich für den ViTEIP-V2a zum EKF Nullkorrelationen; für den TEMINT ergaben sich zum EKF entweder Null- oder negativ-signifikante Korrelationen. Die Korrelation zwischen den Emotionale-Intelligenz-Tests und dem Situationsbezogenen Empathiefragebogen fiel nicht signifikant aus. Insgesamt kann die Korrelation zu Intelligenz als Beleg gesehen werden, dass der ViTEIP-V2 eine kognitive Komponente erfasst; aufgrund der überwiegend geringen oder nicht signifikanten Korrelationen zu anderen Verfahren ist die konvergente Validität jedoch eher kritisch einzuschätzen.

Tabelle 9: Korrelation des ViTEIP-V2a und TEMINT (unterschiedliche Scorings) mit anderen Verfahren bzw. demographischen Variablen

	ViTEIP-V2a				TEMINT			
	Target-Übereinst.	Target-Treffer	Konsens-Anteil	Konsens-Modal	Target-Übereinst.	Target-Treffer	Konsens-Anteil	Konsens-Modal
Stichprobe 1								
IST_Verbal	0.30**	0.26**	0.24**	0.22**	-0.06	-0.03	0.04	0.05
IST_Figural	0.13	0.07	0.05	0.01	-0.07	-0.06	-0.05	-0.03
IST_Merkfähigkeit	0.24**	0.15*	0.12	0.05	-0.09	-0.09	-0.09	-0.06
EKF_Gesamt	0.11	0.22**	0.02	0.09	-0.02	0.01	0.00	0.11
EKF_EE	0.12	0.22**	-0.01	0.06	-0.02	-0.01	0.05	0.12
EKF_EA	0.08	0.12	0.16*	0.22**	0.03	0.04	0.12	0.08
EKF_RG	0.13	0.18*	0.05	0.03	-0.08	-0.08	-0.15	0.13
EKF_EX	0.02	0.10	-0.06	-0.01	0.03	0.06	0.05	0.02
NEO-FFI_Neutrozismus	-0.15	-0.13	-0.08	-0.04	-0.02	0.01	-0.05	-0.15
NEO-FFI_Extraversion	-0.14	-0.06	-0.04	0.00	0.08	0.03	0.11	-0.14
NEO-FFI_Offenheit für Neues	0.03	0.14	-0.01	0.02	-0.13	-0.09	-0.10	0.03
NEO-FFI_Verträglichkeit	-0.03	0.04	0.11	0.17*	-0.05	-0.04	0.03	-0.03
NEO-FFI_Gewissenhaftigkeit	0.15	0.20*	0.06	0.08	0.19*	0.18*	0.14	0.15
<i>Geschlecht¹</i>	0.17*	0.18*	0.10	0.13	-0.04	-0.01	-0.03	0.17*
<i>Alter²</i>	-0.08	0.02	-0.02	0.03	-0.08	-0.08	-0.06	-0.08
<i>Höchster Schulabschluss</i>	0.18*	0.14	0.16*	0.13	-0.05	-0.01	0.02	0.18*
<i>Erfahrungen im Pflegebereich²</i>	0.04	0.11	0.02	0.04	-0.02	-0.04	0.01	0.04
Stichprobe 2a								
WIT	-0.09	-0.10	-0.06	-0.07	0.20*	0.17*	0.10	0.11
Empathiefragebogen	0.02	0.03	-0.10	-0.14	0.01	-0.02	-0.07	-0.05
EKF_Gesamt	-0.08	-0.06	-0.02	-0.04	-0.15*	-0.13	-0.12	-0.08
EKF_EE	0.01	-0.08	0.06	-0.03	-0.12	-0.12	-0.11	-0.07
EKF_EA	-0.11	-0.01	-0.08	0.02	-0.13	-0.09	-0.06	-0.04
EKF_RG	-0.06	-0.03	0.00	-0.01	-0.21**	-0.21**	-0.13	-0.11
EKF_EX	-0.06	-0.05	-0.05	-0.09	0.05	0.06	-0.05	0.01
<i>Geschlecht¹</i>	0.01	0.13	0.14*	0.23**	0.00	0.00	-0.03	-0.03
<i>Alter²</i>	-0.09	0.05	-0.07	0.05	-0.06	-0.01	-0.12	-0.09
<i>Ausbildungsjahrgang</i>	-0.01	0.05	-0.09	-0.03	-0.08	-0.05	-0.07	-0.04
<i>Höchster Schulabschluss</i>	0.10	0.12	-0.12	-0.09	-0.03	0.00	-0.01	0.00
<i>Erfahrungen im Pflegebereich²</i>	-0.01	0.02	-0.02	0.00	-0.02	0.01	0.03	0.05

Anmerkungen:Korrelation nach Pearson. Einseitige Testung, * $p < 0.05$. ** $p < 0.01$.

Für eine detaillierte Beschreibung der Scorings siehe Kapitel 3.3.1.

Für die Berechnungen wurden standardisierte Werte verwendet, siehe Kapitel 3.3.1.1.

IST= Intelligenz-Struktur-Test 2000-R (Amthauer, Brocke, Liepmann, & Beauducel, 2001). EKF= Emotionale-Kompetenz-Fragebogen (Rindermann, 2007). EKF_EE= EKF-Skala Erkennen eigener Emotionen; EKF_EA= EKF-Skala Erkennen Emotionen anderer; EKF_RG= EKF-Skala Regulation und Kontrolle; EKF_EX= EKF-Skala Emotionale Expressivität. NEO-FFI= NEO-Fünf-Faktoren Inventar nach Costa und McCrae (Borkenau & Ostendorf, 1993). WIT= Der Wilde-Intelligenztest (Kersting, Althoff, & Jäger, 2008).

¹ Kodierung: 0= männlich, 1= weiblich.² in Jahren.

Diskriminante Validität. Hierzu zählen die Zusammenhänge zwischen ViTEIP-V2a bzw. TEMINT und den Skalen des NEO-FFI (nur Stichprobe 1) sowie Merkfähigkeit. Diese Korrelationen waren überwiegend nicht signifikant. Bei der Skala Gewissenhaftigkeit ergaben sich über beide Emotionale-Intelligenz-Tests drei positiv-signifikante Korrelationen zu targetbasierten Scorings mit einer maximalen Höhe von $r \leq 0.20^*$. Maximal moderate signifikante Korrelationen zu Merkfähigkeit

fanden sich bei den Targetscorings in Stichprobe 1 für den ViTEIP-V2 und in Stichprobe 2a für den TEMINT. Beide Tests zeigen insgesamt diskriminante Validität

Prüfung der Unterschiedlichkeit der Korrelationen. Es wurde geprüft, inwiefern sich bei den verschiedenen Scorings signifikante Unterschiede in den Korrelationen zwischen ViTEIP-V2 bzw. TEMINT und Kennwerten der anderen Verfahren zeigen. Die Berechnung der Prüfgröße z erfolgte nach Bortz (2005, S. 222-223) mit der Formel für die Prüfung der Unterschiedlichkeit zweier Korrelationskoeffizienten innerhalb einer Stichprobe (*Formel 9, Anhang F*). Nur bei sehr wenigen Korrelationen überschritt die Prüfgröße z den kritischen Prüfwert von $z_{krit}=\pm 1.96$. Insgesamt ergaben sich wie in *Tabelle 10* zu sehen ist, in Stichprobe 1 acht signifikante Unterschiede für den ViTEIP-v2 und sechs für den TEMINT. Alle anderen Korrelationen unterschieden sich nicht signifikant. In der vorliegenden Untersuchung sind die Auswirkungen auf die Korrelationshöhe, die sich durch die Anwendung eines anderen Scorings für die Validität ergeben, minimal.

Tabelle 10: Korrelationen, die sich bei verschiedenen Scorings signifikant unterscheiden

	Merk- fähigk.	EKF	EKF_ EE	EKF_ RG	NEO- FFI_N	NEO- FFI_E	NEO- FFI_O	NEO- FFI_V
ViTEIP Target-Übereinst. – Target-Treffer	x	x	x				x	
ViTEIP Target-Übereinst. – Konsens-Modal	x							x
ViTEIP Target-Treffer – Konsens-Anteil		x	x					
TEMINT Target-Übereinst. – Konsens-Modal				x		x	x	
TEMINT Target-Treffer – Konsens-Anteil				x	x	x		

Anmerkungen:

EKF= Emotionale-Kompetenz-Fragebogen (Rindermann, 2007). EKF_EE= EKF-Skala Erkennen eigener Emotionen; EKF_RG= EKF-Skala Regulation und Kontrolle; NEO-FFI= NEO-Fünf-Faktoren Inventar nach Costa und McCrae (Borkenau & Ostendorf, 1993).

x= signifikante Unterschiede zwischen den Korrelationen.

4.2 Entwicklung von Scorings mit biasfreien Kennwerten

Ausgangspunkt für die nachfolgenden Analysen war das unerwartete Korrelationsmuster zwischen ViTEIP-V2 und TEMINT. Wie beschrieben, zeigten sich bei konsensbasierten Scorings zwischen den Emotionale-Intelligenz-Tests positiv-signifikante Korrelationen. Dies entspricht im Rahmen der konvergenten Validität den Erwartungen. Ein anderes Bild ergab sich bei Target-Übereinstimmung und Target-Treffer: die Korrelationen zwischen den Tests lagen hier erwartungsdiskonform im negativen Bereich.

Grundsätzlich ist es plausibel, dass unterschiedliche Scorings zu unterschiedlichen Korrelationsmustern führen – inwiefern dies der Fall ist, soll bekanntermaßen im Rahmen der vorliegenden Arbeit erforscht werden. Da sich die Veränderung der Korrelationen jedoch ausschließlich auf den Zusammenhang der beiden Emotionale-Intelligenz-Tests untereinander bezieht, und nicht (bzw. nur

in sehr geringem Maße, siehe *Tabellen 9 und 10*) auf die Korrelationen zu anderen Verfahren, ist eine genauere Analyse angebracht.

Identifikation eines Antwortbias auf Seiten der Probanden. Bei Likertskalen kann es vorkommen, dass Probanden nicht die gesamte Antwortskala nutzen sondern ein Antwortverhalten zeigen, das bestimmte Skalenbereiche bevorzugt, z.B. Zustimmungs- oder Ablehnungstendenzen, Tendenzen zur mittleren Antwortkategorie oder sozial erwünschte Antworten (Greenleaf, 1992) (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Um zu prüfen, ob in der vorliegenden Arbeit ein solcher Antwortbias vorlag, wurde für jede Versuchsperson ausgezählt, wie häufig sie in ViTEIP-V2 und TEMINT jeweils die verschiedenen Antwortoptionen wählte²². Pro Person ergaben sich damit entsprechend der Anzahl der Antwortkategorien vier Kennwerte zum Antwortverhalten im ViTEIP und drei Kennwerte im TEMINT; d.h. ein Kennwert, der die Häufigkeit angab, mit der die Antwortoption „Emotion nicht vorhanden“ gewählt wurde, ein Kennwert für die Häufigkeit der Nutzung der Antwort „Emotion schwach vorhanden“ usw. Im nächsten Schritt wurden pro Verfahren die Korrelationen zwischen den Gesamttestwerten bei unterschiedlichen Scorings und den „Kennwerten zum Antwortverhalten“ berechnet. Aus Gründen der Übersichtlichkeit sind in *Tabelle 11* nur die Ergebnisse der Stichprobe 1 aufgeführt; für Stichprobe 2a und 2b ist die *Tabelle A-24 in Anhang G* zu finden.

Für den ViTEIP-V2 zeigte sich, dass Probanden, die häufig „Emotion nicht vorhanden“ ankreuzten, bei allen Auswertungsprozeduren ein besseres Gesamtergebnis erhielten. Entgegengesetzt verhielt es sich beim TEMINT: Hier erhielten die Probanden, die häufig „Emotion stark bis sehr stark vorhanden“ ankreuzten, ein besseres Gesamtergebnis. Besonders auffällig war dies bei den beiden Auswertungen Target-Übereinstimmung und –Treffer, wo die Korrelationen über $r \geq 0.90^{**}$ lagen. Wurde das Antwortverhalten der Probanden über ViTEIP-V2 und TEMINT hinweg betrachtet, fiel eine positive Korrelation auf zwischen der Häufigkeit, mit der in beiden Verfahren vergleichbare Antwortoptionen gewählt wurden (*Tabelle 12* für Stichprobe 1; *Tabelle A-25 in Anhang G* für Stichproben 2a und 2b). Exemplarisch sei hier die Korrelation von $r=0.53^{**}$ zwischen der Häufigkeit, mit der Probanden die Antwortoption „Emotion nicht vorhanden“ im ViTEIP-V2 wählten und der Häufigkeit, mit der Probanden diese Antwortoption im TEMINT wählten, genannt. Dies wird als Hinweis auf die Existenz von Antwortpräferenzen (Antwortbias) seitens der Probanden verstanden. Eine Überprüfung ergab, dass sich diese Antworttendenzen nicht in den anderen eingesetzten Verfahren zeigen, siehe dazu Kapitel 2.1.1, Fußnote 12.

²² Häufigkeit, mit der die Probanden im ViTEIP bzw. TEMINT die jeweiligen Antwortoptionen gewählt haben. Beispiel: Für den Kennwert „Häufigkeit Antwortoption: Emotion nicht vorhanden“ wird ausgezählt, wie häufig jeder Proband innerhalb des Tests die Antwort „Emotion nicht vorhanden“ angekreuzt hat – unabhängig davon, mit welchem Scoring der Test ausgewertet wird.

Ein Schriftwechsel mit dem Autor des MEIS bzw. MSCEIT John Mayer (Mayer, persönliche Korrespondenz im Juli 2011) zum Thema Targetscoring erbrachte keine differenzierten Informationen darüber, ob sich für die Forschergruppe ein ähnliches Bild im Rahmen der Testentwicklung von MEIS und MSCEIT ergab. Follesdal und Hagtvét (2009) berichten jedoch über Items im MSCEIT, die eine negative Korrelation aufweisen zu anderen Items oder Werten, die dasselbe Konstrukt messen sollen. Diese Items würden auch nicht in die Reliabilitätsschätzungen und Berechnungen der Skalenscores einfließen (Follesdal & Hagtvét, 2009, p. 95). Von diesem Vorgehen soll in der vorliegenden Arbeit abgesehen werden. Im Rahmen der *Hypothesen 1c-e* werden targetbasierte Scorings entwickelt, mit dem Ziel den Einfluss des Antwortbias auf Testkennwerte zu minimieren.

Tabelle 11: *Korrelation der Kennwerte zum Antwortverhalten mit den Gesamtestwerten bei unterschiedlichen Scorings (ViTEIP-V2 bzw. TEMINT)*

Stichprobe	Test	Kennwert „Häufigkeit Antwortoption...	Gesamtestwert ermittelt anhand Scoring			
			Target-Übereinst.	Target-Treffer	Konsens-Anteil	Konsens-Modal
1	ViTEIP-V2	...Emotion nicht vorhanden“	0.68**	0.72**	0.57**	0.63**
		...Emotion schwach vorhanden“	0.11	-0.25**	-0.17	-0.30**
		...Emotion mittel vorhanden“	-0.50**	-0.55**	-0.47**	-0.39**
		...Emotion stark vorhanden“	-0.55**	-0.27**	-0.22*	-0.25**
	TEMINT	...Emotion nicht oder sehr schwach vorhanden“	-0.77**	-0.63**	-0.07	-0.24*
		...Emotion schwach bis mittel vorhanden“	-0.31**	-0.50**	-0.56**	-0.44**
		...Emotion stark bis sehr stark vorhanden“	0.90**	0.94**	0.52**	0.56**

Anmerkungen:

Aus Gründen der Übersichtlichkeit finden sich die Ergebnisse der Stichproben 2a und 2b im Anhang.

Zweiseitige Testung, * $p < 0.05$. ** $p < 0.01$.

Für eine detaillierte Beschreibung der Scorings siehe Kapitel 3.3.1.

Kennwert „Häufigkeit Antwortoption...“ = Häufigkeit, mit der Probanden im ViTEIP bzw. TEMINT die jeweilige Antwortoption wählten.

Tabelle 12: *Korrelation der Kennwerte zum Antwortverhalten (ViTEIP-V2 und TEMINT, Stichprobe 1)*

		TEMINT Kennwert „Häufigkeit Antwortoption...		
		...Emotion nicht/sehr schwach vorhanden“	...Emotion schwach bis mittel vorhanden“	...Emotion stark bis sehr stark vorhanden“
ViTEIP Kennwert „Häufigkeit Antwortoption...	„Emotion nicht vorhanden“	0.53**	-0.27**	-0.23**
	„Emotion schwach vorhanden“	-0.02	0.35**	-0.27**
	„Emotion mittel vorhanden“	-0.41**	0.22*	0.16*
	„Emotion stark vorhanden“	-0.32**	-0.14	0.38**

Anmerkungen:

Aus Gründen der Übersichtlichkeit finden sich die Ergebnisse der Stichproben 2a und 2b im Anhang.

Zweiseitige Testung, * $p < 0.05$. ** $p < 0.01$.

fettgedruckt sind die Korrelationen bei vergleichbaren Antwortkategorien der beiden Tests.

Kennwert „Häufigkeit Antwortoption...“ = Häufigkeit, mit der Probanden im ViTEIP bzw. TEMINT die jeweilige Antwortoption wählten.

4.2.1 Profilauswertung

Wie unter 2.1.1.1 beschrieben, wurde für jede Situation²³ berechnet, inwiefern das Profil der Probandenantworten mit dem Profil der Angaben der Originalperson übereinstimmen. Ein hoher positiver Rangkorrelationswert bedeutet, dass das Antwortmuster eines Probanden eine hohe Ähnlichkeit mit dem Antwortmuster der Originalperson besitzt, dieses Muster sich jedoch in der absoluten Ausprägungshöhe unterscheiden kann. Aus den insgesamt zwölf Rangkorrelationen der Situationen wurden für ViTEIP-V2 und TEMINT je ein „Profil-Gesamtwert“ aufsummiert.

Tabelle 13: Deskriptive Statistiken sowie Mittelwertvergleiche zwischen den Stichproben bei Profilauswertung (ViTEIP-V2 und TEMINT)

Verfahren	Stich-probe	M	s ²	α	Kolmogorov-Smirnov	Mittelwertvergleiche zwischen Stichproben					
						1 und 2a		1 und 2b		2a und 2b	
						t	d	t	d	t	d
ViTEIP-V2	1	6.17	1.39	0.45 (0.87)	0.09 (p=0.04)						
	2a	5.38	1.19	0.31 (0.78)	0.04 (p=0.20)	5.49*	0.07	4.31*	0.05	-1.43	-0.02
	2b	5.56	0.94	0.12 (0.53)	0.06 (p=0.20)						
TEMINT	1	9.79	0.30	0.47 (0.86)	0.12 (p=0.00)						
	2a	9.54	0.46	0.36 (0.80)	0.06 (p=0.20)	3.23* ¹	0.02	4.15*	0.03	1.01	-0.01
	2b	9.45	0.46	0.65 (0.93)	0.11 (p=0.00)						

Anmerkungen:

M= Mittelwert; s²= Varianz; α = interne Konsistenz Cronbachs α , Berechnung über Situationen; in Klammern angegeben ist die Reliabilität nach Spearman-Brown-Korrektur zur Testverlängerung auf 99 bzw. 85 Items. Kolmogorov-Smirnov= Test zur Prüfung auf Normalverteilung, mit Signifikanzkorrektur nach Lilliefors, Signifikanzniveau $p \leq 0.01$. Mittelwertvergleiche zwischen Stichproben: t= t-Test für unabhängige Stichproben, zweiseitige Testung zwischen je zwei Stichproben; d= Effektstärke.

¹ Levene-Test der Varianzgleichheit auf dem 0.05 Niveau signifikant.

*Auf dem Niveau 0.05 signifikant.

4.2.1.1 Deskriptive Statistiken

Die deskriptiven Kennwerte bei Anwendung der Profilauswertung sind für ViTEIP-V2 und TEMINT in *Tabelle 13* aufgeführt. Die Mittelwertvergleiche zwischen Stichprobe 1 und den Teilstichproben 2a bzw. 2b fielen für beide Tests signifikant aus, jedoch mit sehr geringen Effektstärken. Die Reliabilitäten waren mit Werten zwischen $\alpha=0.31$ und $\alpha=0.65$ insgesamt eher niedrig; in Stichprobe 2b mit $\alpha=0.12$ (ViTEIP-V2) sogar noch einmal deutlich geringer als in den anderen Stichproben. Bei der Reliabilität ist zu beachten, dass aufgrund der Profilbildung über Situationen die eingerechnete Itemanzahl auf 12 zurückging. Zur besseren Vergleichbarkeit mit den ‚klassischen‘ Scorings (Target-Übereinstimmung, Target-Treffer, Konsens-Anteil, Konsens-Modal) sind in *Tabelle 13* in Klammern

²³ Situationen bestehen aus mehreren Items und bilden inhaltlich zusammengehörige Cluster von Emotionen, so wie sie von den Originalpersonen geschildert wurden.

die internen Konsistenzen bei einer Testverlängerung auf die ursprüngliche Itemanzahl von 99 für den ViTEIP-V2 bzw. 85 für den TEMINT angegeben. Dies stellt jedoch eine rein theoretische Information dar, da für die praktische Anwendung eine solch extreme Testverlängerung aus Gründen der Ökonomie und des zeitlichen Aufwandes bei der Testbearbeitung nicht in Frage käme.

4.2.1.2 Überprüfung der Kennwerte in Bezug auf den Antwortbias

Kann durch die Profilauswertung erreicht werden, dass es in ViTEIP-v2 und TEMINT eine geringere Bevor- oder Benachteiligungen aufgrund des individuellen Ankreuzverhaltens gibt? Um dies zu prüfen, wurden die Korrelationen zwischen dem „Profil-Gesamtwert“ und den „Kennwerten zum Antwortverhalten“ berechnet. Für die Prüfung der Unterschiedlichkeit zweier Korrelationen innerhalb einer Stichprobe wurde nach der Formel von Bortz (2005, S. 222-223) eine Prüfgröße z berechnet (siehe *Formel 9 Anhang F*). Ein Überschreiten des kritischen z -Wertes bedeutet, dass die Korrelation zwischen dem „Kennwert zum Antwortverhalten“ und „Profil-Gesamtwert“ sich signifikant von der Korrelation zwischen dem „Kennwert zum Antwortverhalten“ und „Target-Übereinstimmung“ unterscheidet. Signifikante Unterschiede (unabhängig von der Richtung der Veränderung) sind in der *Tabelle 14* fett gedruckt. Grau hinterlegt sind Korrelationen, die einen erwartungskonformen, geringeren Koeffizienten aufweisen als die Korrelation „Antwortverhalten und Target-Übereinstimmung“.

Tabelle 14: Korrelation „Profil-Gesamtwert“ mit Kennwerten zum Antwortverhalten von ViTEIP-V2 und TEMINT

Test	Kennwert „Häufigkeit Antwortoption...“	„Profil-Gesamtwert“		
		Stichprobe 1	Stichprobe 2a	Stichprobe 2b
ViTEIP-V2	...Emotion nicht vorhanden“	0.34**	0.10	0.17
	...Emotion schwach vorhanden“	-0.16	0.21*	-0.12
	...Emotion mittel vorhanden“	-0.27**	-0.01	-0.10
	...Emotion stark vorhanden“	-0.07	-0.18*	-0.11
TEMINT	...Emotion nicht/sehr schwach vorhanden“	-0.53**	-0.38**	-0.61**
	...Emotion schwach bis mittel vorhanden“	-0.23*	-0.12	0.16
	...Emotion stark bis sehr stark vorhanden“	0.64**	0.39**	0.50**

Anmerkungen:

Korrelation nach Pearson. Einseitige Testung, * $p < 0.05$. ** $p < 0.01$.

Profil-Gesamtwert= Kennwert aus der Profilauswertung. Häufigkeit Antwortoption= Häufigkeit, mit der die Probanden im ViTEIP bzw. TEMINT die jeweiligen Antwortoptionen gewählt haben.

Signifikante Unterschiede (unabhängig von der Richtung der Veränderung) in der Korrelation „Antwortverhalten und Profilauswertung“ und zu der Korrelation „Antwortverhalten und Target-Übereinstimmung“ sind **fett** gedruckt; **grau hinterlegt**= signifikanten Veränderung in die erwartete Richtung.

Die überwiegend geringeren Korrelationen zwischen den ViTEIP-Kennwerten zum Antwortverhalten und dem Gesamtwert zeigen, dass das Antwortverhalten bei der Profilauswertung einen deutlich geringeren Einfluss auf den Gesamtwert hatte als bei der Auswertung mittels Target-Übereinstimmung. Auch beim TEMINT reduzierten sich die Korrelationen zum Antwortverhalten bei Profilauswertung signifikant. Die Höhe der Korrelation war an einigen Stellen nach wie vor hoch und weist damit auf einen Zusammenhang zwischen der Häufigkeit der gewählten Antwortoption und dem „Profil-Gesamtwert“ hin. Insgesamt wurden die erhofften Veränderungen erzielt, wenn auch für den TEMINT etwas weniger deutlich als für den ViTEIP-V2.

4.2.1.3 Korrelationsanalysen

Sowohl beim ViTEIP-V2 als auch beim TEMINT lagen innerhalb des jeweiligen Tests zwischen Target-Übereinstimmung und Profil-Gesamtwert hohe Konvergenzen mit Korrelationen zwischen $r \geq 0.63^{**}$ und $r \leq 0.81^{**}$ vor (Tabelle 15). Der Zusammenhang zwischen den beiden Emotionale-Intelligenz-Tests verschiebt sich von der negativen Korrelation bei Target-Übereinstimmung (zwischen $r \geq -0.33^{**}$ und $r \leq -0.06$ n.s.) hin zu schwach positiven Korrelationen (zwischen $r \geq 0.09$ n.s. und $r \leq 0.16^*$) bei Profil-Gesamtwert. Eine Tabelle mit Korrelationen zwischen „Profil-Gesamtwert“ und den anderen eingesetzten Verfahren ist in *Anhang B* eingefügt. Die Berechnung von minderungskorrigierten Korrelationen²⁴ ist nicht möglich, da – wie unter 4.1.1 aufgeführt – bei den vorliegenden Daten die Voraussetzung der Eindimensionalität nicht gegeben ist (Bühner, 2006).

Tabelle 15: Korrelationen Target-Übereinstimmung und Profil-Gesamtwert (ViTEIP-V2 und TEMINT)

Stichprobe	Verfahren	Scoring	ViTEIP-V2		TEMINT	
			Target-Übereinstimmung	Profil-Gesamtwert	Target-Übereinstimmung	Profil-Gesamtwert
1	ViTEIP-V2	Target-Übereinstimmung				
		Profil-Gesamtwert	0.75**			
	TEMINT	Target-Übereinstimmung	-0.26**	-0.02		
		Profil-Gesamtwert	-0.09	0.09	0.80**	
2a	ViTEIP-V2	Target-Übereinstimmung				
		Profil-Gesamtwert	0.63**			
	TEMINT	Target-Übereinstimmung	-0.33**	0.02		
		Profil-Gesamtwert	-0.08	0.16*	0.78**	
2b	ViTEIP-V2	Target-Übereinstimmung				
		Profil-Gesamtwert	0.72**			
	TEMINT	Target-Übereinstimmung	-0.06	0.03		
		Profil-Gesamtwert	0.03	0.11	0.81**	

Anmerkungen:

Korrelation nach Pearson. Einseitige Testung, * $p < 0.05$, ** $p < 0.01$.

Target-Übereinstimmung= Für eine Beschreibung siehe Kapitel 3.3.1. Profil-Gesamtwert= Kennwert aus der Profilauswertung.

²⁴ Durch die doppelte Minderungskorrektur kann berechnet werden, wie hoch zwei Variablen korrelieren, wenn beide Variablen perfekt reliabel (d.h. ohne Messfehler) erfasst würden.

4.2.2 Übertragung der Signalentdeckungstheorie auf Emotionale-Intelligenz-Tests

Um die grundsätzliche Übertragbarkeit der Signalentdeckungstheorie auf Emotionale Intelligenz zu überprüfen, wurden – wie unter 2.1.1.2 beschrieben – zwei Annahmen bezüglich der Verteilungen der Rausch- und Signalkurve ungeprüft vorausgesetzt (vgl. (Wickens, 2002)). Sind diese Annahmen erfüllt, kann die Detektionsleistung (\hat{d}') als Abstand zwischen den beiden Verteilungen und $\hat{\beta}$ als Modellparameter, der das Entscheidungsverhalten ausdrückt, interpretiert werden. Sollte sich die Signalentdeckungstheorie prinzipiell als Auswertungsstrategie für Emotionale-Intelligenz-Tests eignen, sind Studien über die Gültigkeit dieser Voraussetzungen erforderlich.

Tabelle 16: Kennwerte aus der Signalentdeckungstheorie in der Übertragung auf Emotionale-Intelligenz-Tests

Kennwert	Berechnung	
Identifikation	$p(\text{Identifikation}) = \frac{\text{Anzahl der Treffer}}{\text{Gesamtanzahl Nicht-Null-Items}}$	(Formel 2)
Verpasser (Fehler zweiter Art)	$p(\text{Verpasser}) = \frac{\text{Anzahl der Verpasser}}{\text{Gesamtanzahl Nicht-Null-Items}}$	(Formel 3)
Falscher Alarm (Fehler erster Art)	$p(\text{falscher Alarm}) = \frac{\text{Anzahl der falschen Alarme}}{\text{Gesamtanzahl Null-Items}}$	(Formel 4)
Korrekte Zurückweisung	$p(\text{korr. Zurückweisung}) = \frac{\text{Anzahl der korr. Zurückweisung}}{\text{Gesamtanzahl Null-Items}}$	(Formel 5)
Detektionsleistung \hat{d}'	$\hat{d}' = Z(\text{Identifikation}) - Z(\text{falscher Alarm})$	(Formel 6)
Reaktionsneigung \hat{c}	$\hat{c} = -0.5 * (Z(\text{falscher Alarm}) + Z(\text{Identifikation}))$	(Formel 7)
Emotionale Sensitivität	Beschreibung siehe Text („Zwei weitere neue Kennwerte“)	
Emotionale Spezifität	Beschreibung siehe Text („Zwei weitere neue Kennwerte“)	

Anmerkungen:

Nicht-Null-Items= Items, bei denen die Antwort der Originalperson „Emotion schwach/mittel/stark vorhanden“ lautet; Null-Items= Items, bei denen die Antwort der Originalperson „Emotion nicht vorhanden“ lautet.

Treffer= die Antwort des Probanden lautet übereinstimmend mit der Originalperson „Emotion schwach/mittel/stark vorhanden“; Verpasser= die Antwort des Probanden lautet fälschlicherweise „Emotion nicht vorhanden“; Falscher Alarm= die Antwort des Probanden lautet fälschlicherweise „Emotion schwach/mittel/stark vorhanden“; Korrekte Zurückweisung= die Antwort des Probanden lautet übereinstimmend mit der Originalperson „Emotion nicht vorhanden“.

Z= Verteilungsfunktion der Standardnormalverteilung. Formeln nach Wickens (2002).

Berechnung der Kennwerte. Für die Auswertung der Emotionale-Intelligenz-Tests mittels Signalentdeckungstheorie wurde zunächst eine Unterteilung in Null- und Nicht-Null-Items vorgenommen. Null-Items sind Items, bei denen die Originalpersonen angaben, die entsprechende Emotion sei nicht vorhanden. Nicht-Null-Items sind Items, bei denen die Angabe der Originalperson lautete „Emotion vorhanden“. In der Übertragung aus der Signalentdeckungstheorie, bei der es in den Experimenten Durchgänge mit „Rauschen“ und Durchgänge mit „Signal“ gibt, wurden die Null-Items zu „Rauschen“ (d.h. keine Reiz vorhanden) und die Items, in denen eine Emotion (d.h. ein

Reiz) vorhanden ist, zu „Signal“ subsummiert. Zudem wurden die Angaben der Probanden betrachtet: Ist die Antwort des Probanden „Emotion nicht vorhanden“ oder „Emotion vorhanden“?

Aus der Kombination der Angaben der Originalpersonen und den Angaben der Probanden ergeben sich die in *Tabelle 16* aufgeführten Kennwerte. „Identifikation“ und „Korrekte Zurückweisung“ beschreiben die Situation, bei der Probanden in ihren Antworten mit den Angaben der Originalperson übereinstimmen: Identifikation bezieht sich auf „Emotion vorhanden“ und korrekte Zurückweisung auf „Emotion nicht vorhanden“. Wird in Items, bei denen eine Originalperson die Anwesenheit der spezifischen Emotion konstatiert hat, dies von Probanden nicht erkannt, handelt es sich um „Verpasser“. Tritt die entgegengesetzte Situation ein und Probanden vermuten fälschlicherweise das Vorhandensein einer Emotion, handelt es sich um „Falschen Alarm“.

Zur Berechnung der Schätzer für die Detektionsleistung \hat{d}' und der Reaktionsneigung \hat{c} wurden die relativen Häufigkeiten der Treffer und Falschen Alarme ermittelt, anschließend mit Excel die Unterschreitungswahrscheinlichkeiten der Standardnormalverteilung berechnet und diese Werte in die *Formeln 6 bzw. 7* eingesetzt. Es gab Probanden, die im TEMINT keinen Falschen Alarm anzeigten und dementsprechend dieser Wert bei Null lag. Um dennoch einen Z-Wert für die Berechnung der Detektionsleistung und Reaktionsneigung zu erhalten, wurden behelfsweise die kleinstmöglichen Werte ($p(fA)=0.001$) aus der Normalverteilungstabelle (Wickens, 2002, p. 250) eingesetzt²⁵.

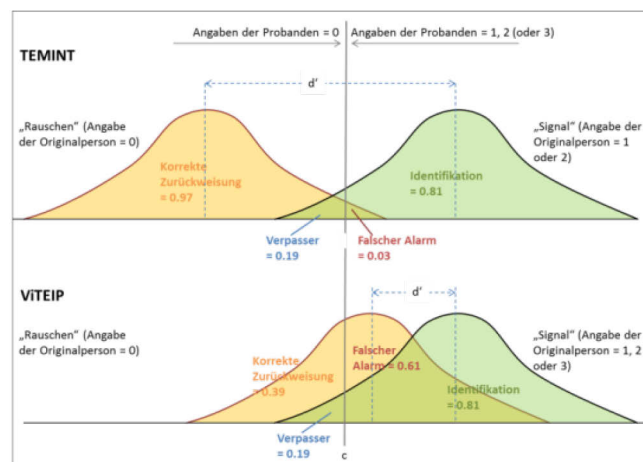
Ein weiteres Scoring mit zwei neuen Kennwerten. Da bei den Kennwerten, die direkt aus der Signalentdeckungstheorie (SET) übertragen wurden, die Emotionsintensität keine Berücksichtigung findet, wurden zusätzlich explorativ ein neues Scoring untersucht, das eine Kombination von Target-Übereinstimmung und Signalentdeckungstheorie darstellt. Für jedes Item wurde zunächst der Übereinstimmungswert mit dem Target berechnet. Je weiter die Antwort des Probanden von der Angabe der Originalperson entfernt liegt, desto geringer ist der Übereinstimmungswert; antwortet ein Proband in Übereinstimmung mit der Antwort der Originalperson wurde dies mit dem Wert „3“ kodiert; antwortet der Proband um eine Einheit abweichend nach rechts oder links, ergibt dies einen Übereinstimmungswert von „2“ usw. (siehe Kapitel 3.3.1). Anschließend wurden alle Items, die von Originalpersonen mit „Emotion vorhanden“ kommentiert wurden, zu einem Gesamtwert „Emotionale Sensitivität“ aufsummiert, ebenso alle Items mit der Angabe der Originalperson „Emotion nicht vorhanden“ zum Gesamtwert „Emotionale Spezifität“.

²⁵ Der Wert ($p(fA)=0.001$) entspricht dem Bereich von drei Standardabweichungen rechts bzw. links des Erwartungswertes. In dieser Fläche liegen 99.73% aller Werte einer normalverteilten Zufallsvariable. Als Alternative zu der Ersetzung der Werte wäre ein mögliches Vorgehen gewesen, alle Probanden mit einer Quote von 100% bei „Korrekte Zurückweisung“ aus den Berechnungen auszuschließen. Die Stichproben würden sich dadurch auf $N_{Stichprobe_1}=39$, $N_{Stichprobe_2a}=83$ und $N_{Stichprobe_2b}=62$ deutlich minimieren. Dieses Vorgehen wird von der Autorin als nicht sinnvoll erachtet.

Der Grundgedanke der SET mit der Unterscheidung von „Rauschen“ (d.h. Null-Items) und „Signal“ (d.h. Nicht-Null-Items) bleibt somit erhalten; durch die Verwendung der Übereinstimmungswerte mit den Abstufungen je nach Nähe/Distanz zur richtigen Antwort werden vorhandene Varianzen beibehalten. Zudem findet in dem Wert „Emotionale Sensitivität“ eine Bündelung von Identifikation und Verpasser bzw. bei „Emotionale Spezifität“ eine Bündelung von Korrekter Zurückweisung und Falscher Alarm statt. Anders als bei klassischen Signalentdeckungsexperimenten, bei denen es um eine Ja/Nein-Entscheidung bezüglich des Vorliegens eines Signals geht, soll es hier nicht nur darum gehen ob grundsätzlich eine beliebige Emotion vorliegt oder nicht. Vielmehr ist es von Bedeutung, dass das Vorliegen einer spezifischen Emotion korrekt identifiziert wird; die jeweilige Emotion wird definiert durch die Angabe der Originalperson.

4.2.2.1 Deskriptive Statistiken

Die mittleren Identifikationen (und somit auch die Verpasser) lagen innerhalb der Stichproben für den ViTEIP bzw. TEMINT jeweils in vergleichbaren Größenordnungen (*Tabelle 17*): in Stichprobe 1 (Studenten) lag der Anteilswert der Identifikationen für beide Tests bei 0.81, in Stichprobe 2a (Pflegeschüler) bei 0.83 und in der Stichprobe 2b (Pflegeschüler) für den ViTEIP-V2 bei 0.65 bzw. TEMINT bei 0.79. Deutliche Unterschiede fanden sich in den Reaktionen auf das „Rauschen“ (d.h. die Items, bei denen die Antwort der Originalperson „Emotion nicht vorhanden“ lautet). Beim TEMINT lag der durchschnittliche Wert für korrekte Zurückweisung in allen drei Stichproben über 95%; dementsprechend gering fielen die Werte für falschen Alarm aus. Beim ViTEIP hingegen gelang den Probanden in Stichprobe 1 und 2a im Mittel nur bei gut einem Drittel der Items (39% bzw. 32%) die korrekte Zurückweisung. In Stichprobe 2b erhöhte sich dieser Wert auf 60%.



Anmerkungen:

Kurvenverlauf exemplarisch dargestellt, entspricht nicht der realen Datenverteilung der vorliegenden Untersuchung. Angegebene Zahlen sind Mittelwerte der Kennwerte in Stichprobe 1 (siehe *Tabelle 17*); Angaben als Anteilswerte.

Abbildung 10: Schematische Darstellung der Rausch- und der Signalkurve für TEMINT und ViTEIP

Anhand der Werte von \hat{d}' und der schematischen Darstellung (*Abbildung 10*) ist zu erkennen, dass die Verteilungen von „Rauschen“ und „Signal“ beim TEMINT ($\hat{d}'=3.47$, Stichprobe 1) weit auseinander lagen und es den Probanden im Mittel verhältnismäßig leicht fiel, eine richtige Entscheidung zu fällen. Für den ViTEIP-V2 weisen die hohen Werte in Falscher Alarm und dem geringeren \hat{d}' darauf hin, dass die Verteilungen näher beieinander lagen (*Tabelle 17*).

Beim ViTEIP-V2 unterschieden sich die Stichproben deutlich in den Mittelwerten von Emotionaler Sensitivität und Emotionaler Spezifität: Stichprobe 2b, die das Antwortformat „Zuordnen“ bearbeitete, schnitt bei Emotionaler Sensitivität im Vergleich zu den anderen beiden Gruppen (Antwortformat „Ankreuzen“) am schlechtesten, bei Emotionaler Spezifität am besten ab. Im direkten Vergleich der Stichproben 1 und 2a (identisches Antwortformat) erreichten die Studenten ein besseres Ergebnis als die Pflegeschüler, obwohl diesen der pflegerische Kontext vertrauter sein sollte. Im TEMINT waren die Unterschiede sehr gering. Die internen Konsistenzen lagen für Emotionaler Spezifität bei ViTEIP-V2 und für Emotionaler Sensitivität bei TEMINT über $\alpha=0.80$; bei Emotionaler Sensitivität bei ViTEIP-V2 und Emotionaler Spezifität bei TEMINT im Schnitt bei $\alpha=0.50$.

Tabelle 17: Kennwerte der Signalentdeckungstheorie in der Übertragung auf ViTEIP-V2 und TEMINT

	Kennwert	Stichprobe 1		Stichprobe 2a		Stichprobe 2b	
		<i>M</i>	<i>s</i> ²	<i>M</i>	<i>s</i> ²	<i>M</i>	<i>s</i> ²
ViTEIP-V2	Identifikation	0.81	0.01	0.83	0.01	0.65	0.01
	Verpasser	0.19	0.01	0.17	0.01	0.35	0.01
	Falscher Alarm	0.61	0.04	0.68	0.03	0.40	0.02
	Korrekte Zurückweisung	0.39	0.04	0.32	0.03	0.60	0.02
	Detektionsleistung \hat{d}'	0.63	0.26	0.53	0.13	0.66	0.09
	Reaktionsneigung \hat{c}	-0.65	0.25	-0.80	0.27	-0.07	0.13
	Emotionale Sensitivität	125.56	55.73	122.22	73.80	111.74	72.33
	Emotionale Spezifität	82.15	245.29	72.18	277.45	90.56	199.27
TEMINT	Identifikation	0.81	0.01	0.83	0.01	0.79	0.01
	Verpasser	0.19	0.01	0.17	0.01	0.21	0.01
	Falscher Alarm	0.03	0.00	0.05	0.00	0.04	0.00
	Korrekte Zurückweisung	0.97	0.00	0.95	0.00	0.96	0.00
	Detektionsleistung \hat{d}'	3.47	0.61	3.07	0.67	3.16	0.66
	Reaktionsneigung \hat{c}	0.82	0.20	0.55	0.25	0.70	0.25
	Emotionale Sensitivität	82.64	83.66	84.10	82.40	80.54	118.53
	Emotionale Spezifität	53.16	2.12	52.09	5.38	52.45	3.77

Anmerkungen:

M= Mittelwert; *s*²= Varianz. Beschreibung der Kennwerte siehe Tabelle 16 und Text.

4.2.2.2 Detektionsleistung, Emotionale Sensitivität und Emotionale Spezifität als Leistungsindikatoren

Es wurde geprüft, ob sich die Kennwerte Detektionsleistung, Emotionale Sensitivität und Emotionale Spezifität als Leistungsindikatoren für Emotionale-Intelligenz-Tests eignen. Die

Erwartung war, dass die Fähigkeitsparameter weitestgehend unabhängig vom Antwortverhalten sind. Zudem wurden Korrelationen zu anderen Variablen betrachtet, um Validitätsnachweise zu erbringen.

Antwortverhalten. In *Tabelle 18* sind Korrelationen zwischen Detektionsleistung und Kennwerten des Antwortverhaltens dargestellt. Es wurde überprüft, inwieweit sich die Korrelationen von der Korrelation zwischen Antwortverhalten und Target-Übereinstimmung unterscheiden (siehe *Formel 9 im Anhang F*). Grau hinterlegt sind Korrelationen mit signifikanter Veränderung in die erwartete Richtung, d.h. einem geringeren Zusammenhang zwischen Antwortverhalten und Kennwert.

Tabelle 18: Korrelation der Kennwerte zum Antwortverhalten zu „Detektionsleistung“ (ViTEIP-V2 und TEMINT)

Test	Kennwert „Häufigkeit Antwortoption...	Detektionsleistung \hat{d}'			Reaktionsneigung \hat{c}		
		Stich- probe 1	Stich- probe 2a	Stich- probe 2b	Stich- probe 1	Stich- probe 2a	Stich- probe 2b
VITEIP-V2	...Emotion nicht vorhanden“	0.48**	0.25**	0.38**	0.94**	0.94**	0.99**
	...Emotion schwach vorhanden“	-0.15	0.01	-0.03	-0.18*	-0.17*	-0.32**
	...Emotion mittel vorhanden“	-0.35**	-0.21**	-0.21*	-0.68**	-0.58**	-0.69**
	...Emotion stark vorhanden“	-0.22**	-0.13	-0.37**	-0.55**	-0.41**	-0.77**
TEMINT	...Emotion nicht/sehr schwach vorhanden“	0.03	0.19*	0.03	0.68**	0.75**	0.72**
	...Emotion schwach bis mittel vorhanden“	-0.25*	-0.15*	-0.13	-0.36**	-0.27**	-0.43**
	...Emotion stark bis sehr stark vorhanden“	0.18	-0.02	0.17*	-0.28*	-0.27**	-0.32**

Anmerkungen:

Korrelation nach Pearson. Einseitige Testung, * $p < 0.05$. ** $p < 0.01$.

Häufigkeit Antwortoption= Häufigkeit, mit der die Probanden im ViTEIP die jeweiligen Antwortoptionen gewählt haben.

Signifikante Unterschiede zwischen der Korrelation „Antwortverhalten und Detektionsleistung“ und der Korrelation „Antwortverhalten und Target-Übereinstimmung“ sind **fett** gedruckt; **grau hinterlegt**= signifikanten Veränderung in die erwartete Richtung.

Die Detektionsleistung wies in den drei Stichproben bei ViTEIP-V2 und TEMINT einen überwiegend schwächeren Zusammenhang zum Antwortverhalten der Probanden auf. Die wenigen Korrelationen, bei denen sich keine signifikante Veränderung im Vergleich zu Target-Übereinstimmung zeigte, fielen gering aus, so dass der Einfluss des Antwortverhaltens auf den Gesamtwert nur marginal ist. Zum Vergleich sind in *Tabelle 18* die Korrelationen der Reaktionsneigung mit den Kennwerten des Antwortverhaltens dargestellt. Diese fielen deutlich größer aus – was der Konzeption der Reaktionsneigung als Indikator für Antworttendenzen entspricht. Der Einfluss des Antwortvorhaltens der Probanden auf Emotionale Sensitivität und Emotionale Spezifität wurde für Nicht-Null- und Null-Items getrennt betrachtet. Verglichen wurden die Korrelationen der Kennwerte zum Antwortverhalten und jeweils zu Emotionaler Sensitivität und Emotionaler Spezifität (*Tabelle 19*). Entscheidend war, dass bei Null-Items die Korrelation zwischen Emotionaler Spezifität und „Häufigkeit Antwortoption Emotion nicht vorhanden“ und bei Nicht-Null-Items die Korrelation

zwischen Emotionaler Sensitivität und „Häufigkeit Antwortoption Emotion stark/mittel vorhanden“ signifikant positiv ausfällt. Anhand der grau hinterlegten Zellen ist zu sehen, dass dies – mit Ausnahme in Stichprobe 2a mit einer nicht signifikanten Korrelation zwischen Emotionaler Sensitivität und „Emotion stark vorhanden“ – für den ViTEIP und TEMINT zutrifft.

Insgesamt zeigten sowohl die Detektionsleistung als auch „Emotionale Sensitivität“ und „Emotionale Spezifität“ einen signifikant geringeren Zusammenhang zum Antwortverhalten der Probanden als es bei Target-Übereinstimmung der Fall war.

Tabelle 19: Korrelation der Kennwerte zum Antwortverhalten zu „Emotionale Sensitivität“ bzw. „Emotionale Spezifität“ (ViTEIP-V2 und TEMINT)

Test	Items	Kennwert „Häufigkeit Antwortoption...	Stichprobe 1		Stichprobe 2a		Stichprobe 2b	
			eSens	eSpez	eSens	eSpez	eSens	eSpez
ViTEIP-V2	Nicht-Null	...Emotion nicht vorhanden“	-0.71**	0.67**	-0.69**	0.68**	-0.82**	0.64**
		...Emotion schwach vorhanden“	-0.19*	0.44**	0.17*	0.43**	0.15	0.14
		...Emotion mittel vorhanden“	0.49**	-0.34**	0.50**	-0.02	0.50**	-0.10
		...Emotion stark vorhanden“	0.38**	-0.60**	0.09	-0.69**	0.51**	-0.69**
	Null	...Emotion nicht vorhanden“	-0.33**	0.93**	-0.52**	0.86**	-0.48**	0.94**
		...Emotion schwach vorhanden“	0.16	-0.29**	0.44**	0.14	0.38**	-0.22*
		...Emotion mittel vorhanden“	0.28**	-0.85**	0.41**	-0.61**	0.31**	-0.65**
		...Emotion stark vorhanden“	0.22*	-0.76**	-0.02	-0.81**	0.26**	-0.86**
TEMINT	Nicht-Null	...Emotion nicht/sehr schwach vorhanden“	-0.88**	0.19*	-0.82**	0.21*	-0.89**	-0.01
		...Emotion schwach bis mittel vorhanden“	-0.31**	-0.08	-0.36**	0.12	0.03	0.04
		...Emotion stark bis sehr stark vorhanden“	0.91**	-0.09	0.83**	-0.23**	0.84**	0.05
	Null	...Emotion nicht/sehr schwach vorhanden“	-0.13	0.98**	-0.20*	0.96**	0.03	0.95**
		...Emotion schwach bis mittel vorhanden“	0.13	-0.91**	0.17*	-0.71**	0.13	-0.65**
		...Emotion stark bis sehr stark vorhanden“	-0.01	-0.47**	0.14	-0.72**	-0.03	-0.46**

Anmerkungen:

Korrelation nach Pearson. Einseitige Testung, * $p < 0.05$. ** $p < 0.01$.

Nicht-Null= Items, bei denen die Antwort der Originalperson „Emotion schwach/mittel/stark vorhanden“ lautet; Null= Items, bei denen die Antwort der Originalperson „Emotion nicht vorhanden“ lautet. Kennwert Häufigkeit Antwortoption= Häufigkeit, mit der die Probanden im ViTEIP bzw. TEMINT diese Antwortoptionen gewählt haben. eSens= Emotionale Sensitivität, eSpez= Emotionale Spezifität.

Korrelationen zwischen ViTEIP-V2 und TEMINT. Die Detektionsleistung korrelierte bei ViTEIP-V2 und TEMINT positiv in unterschiedlicher Höhe mit Target-Übereinstimmung; die Werte liegen zwischen $r \geq 0.28^{**}$ und $r \leq 0.82^{**}$ (Tabellen A-28 bis A-30 in Anhang H). Die Korrelation der Reaktionsneigung (als Indikator für Antworttendenzen) mit dem Gesamtübereinstimmungswert war für den ViTEIP positiv, für den TEMINT hingegen negativ. Hier wurde der beschriebene Antwortbias in Form eines unterschiedlichen Einflusses des Antwortverhaltens auf die Gesamttestleistung deutlich. Zudem gab es beim ViTEIP-V2 eine hohe Korrelation zwischen Übereinstimmungswert und Emotionaler Spezifität (zwischen $r \geq 0.81^{**}$ bis $r \leq 0.88^{**}$), beim TEMINT korrelierte der

Übereinstimmungswert sehr hoch mit Emotionaler Sensitivität (zwischen $r \geq 0.91^{**}$ und $r \leq 0.99^{**}$). Die jeweils andere Kategorie korrelierte mit dem Übereinstimmungswert bei beiden Tests nicht oder nur gering (zwischen $r \geq 0.04$ n.s. und $r \leq 0.29^{**}$).

Zwischen ViTEIP-V2 und TEMINT korrelierten die Detektionsleistungen nicht signifikant mit $r=0.05$ (Stichprobe 1 und 2a) und $r=0.07$ (Stichprobe 2b) sowie die Reaktionsneigungen signifikant mit $r=0.39^{**}$ (Stichprobe 1) und $r=0.35^{**}$ (Stichprobe 2a und 2b). Emotionale Sensitivität und Emotionale Spezifität zeigten zwischen den Tests signifikante Zusammenhänge in der Größenordnung von $r \geq 0.18^*$ bis $r \leq 0.38^{**}$. Vor allem das letztgenannte Ergebnis ist von Bedeutung, da hier deutlich wird, dass es – anders als es die Korrelationen zwischen targetbasierten Auswertungen vermuten ließen – sehr wohl Zusammenhänge zwischen der Leistung eines Probanden in ViTEIP-V2 und TEMINT gibt. Durch die Antworttendenzen und die gemeinsame Auswertung aller Items wurden diese Zusammenhänge möglicherweise verdeckt.

Tabelle 20: Korrelation der Kennwerte ViTEIP-V2 und TEMINT mit anderen Verfahren (Stichprobe 1)

Kennwert	ViTEIP-V2			TEMINT		
	\hat{d}'	eSens	eSpez	\hat{d}'	eSens	eSpez
d2-R_Konzentrationsleistung	-0.04	-0.14	-0.02	0.06	-0.12	0.02
IST_Verbal	0.24 ^{**}	-0.01	0.31 ^{**}	0.01	-0.09	0.05
IST_Figural	0.11	-0.02	0.13	-0.01	-0.09	0.00
IST_Merkfähigkeit	0.21 [*]	0.09	0.18 [*]	-0.07	-0.07	-0.02
EKF_Gesamt	0.29 ^{**}	-0.09	0.17 [*]	0.07	-0.06	0.10
EKF_EE	0.26 ^{**}	-0.10	0.18 [*]	0.14	-0.06	0.14
EKF_EA	0.16	-0.19 [*]	0.20 [*]	0.07	0.00	0.06
EKF_RG	0.25 ^{**}	0.10	0.08	-0.02	-0.08	0.03
EKF_EX	0.10	-0.04	0.06	-0.02	0.00	0.01
NEO-FFI_Neurotizismus	-0.17 [*]	-0.05	-0.11	-0.02	0.02	-0.06
NEO-FFI_Extraversion	0.02	-0.13	-0.06	-0.06	0.05	0.02
NEO-FFI_Offenheit für Neues	0.28 ^{**}	-0.06	0.08	-0.14	-0.14	-0.03
NEO-FFI_Verträglichkeit	0.07	-0.16 [*]	0.07	0.19 [*]	-0.10	0.21 [*]
NEO-FFI_Gewissenhaftigkeit	0.21 [*]	0.15	0.05	0.16 [*]	0.13	0.17 [*]
Geschlecht ¹	0.22 ^{**}	-0.09	0.22 [*]	0.03	-0.06	0.10
Alter ²	0.03	0.02	-0.09	-0.03	-0.07	-0.02
Schulabschluss	0.18 [*]	0.03	0.18 [*]	0.01	-0.08	0.02
Erfahrungen im Pflegebereich ²	0.11	0.06	0.01	0.00	-0.05	0.04

Anmerkungen:

Aus Gründen der Übersichtlichkeit finden sich die Ergebnisse der Stichproben 2a und 2b im Anhang.

Korrelation nach Pearson. Einseitige Testung, $*p < 0.05$. $**p < 0.01$.

\hat{d}' = Detektionsleistung, eSens= Emotionale Sensitivität, eSpez= Emotionale Spezifität.

d2-R= Aufmerksamkeits-Konzentrations-Test d2, revidierte Form (Brickenkamp, Schmidt-Atzert, & Liepmann, 2010). IST= Intelligenz-Struktur-Test 2000-R (Amthauer, Brocke, Liepmann, & Beauducel, 2001). EKF= Emotionale-Kompetenz-Fragebogen (Rindermann, 2007). EKF_EE= EKF-Skala Erkennen eigener Emotionen; EKF_EA= EKF-Skala Erkennen Emotionen anderer; EKF_RG= EKF-Skala Regulation und Kontrolle; EKF_EX= EKF-Skala Emotionale Expressivität. NEO-FFI= NEO-Fünf-Faktoren Inventar nach Costa & McCrae (Borkenau & Ostendorf, 1993).

¹ Kodierung: 0= männlich, 1= weiblich.

² in Jahren.

Korrelationen zu anderen Verfahren. Beim ViTEIP-V2 (Stichprobe 1, *Tabelle 20*) korrelierte die Detektionsleistung im Sinne der konvergenten Validität signifikant positiv mit verbaler Intelligenz sowie dem EKF Gesamtwert und den EKF Skalen „Erkennen eigener Emotionen“ und „Regulation und Kontrolle“. Die Korrelationen lagen im geringen bis mittleren Bereich, was für konvergente Zusammenhänge eher als niedrig zu deuten ist. Das Korrelationsmuster zu den BIG-5-Skalen entsprach überwiegend sowohl in der Richtung der Vorzeichen als auch in der Höhe der Korrelationen den Erwartungen zur diskriminanten Validität. Die signifikante Korrelation zu Gewissenhaftigkeit und Merkfähigkeit waren bezüglich der absoluten Höhe betrachtet erwartungskonform, erscheinen jedoch aufgrund der geringen Höhe der konstrukt-nahen Korrelationen vergleichsweise hoch.

Für Emotionale Spezifität zeigte sich ein vergleichbares Bild der Zusammenhänge wie für Detektionsleistung; Emotionale Sensitivität korrelierte überwiegend nicht signifikant. Bei Stichprobe 2a und 2b (*Anhang H, Tabellen A-26 und A-27*) ergab sich ein Bild, das geprägt war von nicht-signifikanten Korrelationen und Korrelationen mit negativem Vorzeichen. Die Korrelationen des TEMINT waren gekennzeichnet durch heterogene Muster über die Stichproben hinweg und fielen sowohl für die konvergente als auch die diskriminante Validität mehrheitlich nicht signifikant aus.

4.3 Antwortformate

In der zweiten Untersuchung bearbeitete die Teilstichprobe 2a den ViTEIP-V2a mit dem Antwortformat „Ankreuzen“ und die Teilstichprobe 2b den ViTEIP-V2b mit dem Antwortformat „Zuordnen“ (*Anhang A, Abbildung A-1 und A-2*). Bei „Ankreuzen“ wurden die Probanden gebeten, auf einer vierstufigen Likertskala die Intensität der aufgeführten Emotionen einzuschätzen. Bei „Zuordnen“ war die Aufgabe der Probanden, die vorgegebenen Emotionen in Bezug auf die Intensität in vier verschiedene Kategorien zu sortieren. Die Zuteilung der Probanden auf die Teilstichproben erfolgte während der Datenerhebung nach dem Zufallsprinzip. Wie in Kapitel 3.1 beschrieben, waren die Teilstichproben hinsichtlich demographischer Variablen vergleichbar. Auch in den Kennwerten der weiteren eingesetzten Verfahren (*Anhang D, Tabelle A-5*) und in Bezug auf die Einschätzung der Akzeptanz des ViTEIP-V2 (*Anhang I, Tabelle A-31*) gab es zwischen den Teilstichproben keine Unterschiede. Um auch die Gesamttestergebnisse (Stichprobenmittelwerte) vergleichen zu können, wurden die Daten ohne z-Transformation verwendet. Neben den ‚klassischen‘ Scorings²⁶ werden

²⁶ ‚klassische‘ Scorings = Target-Übereinstimmung, Target-Treffer, Konsens-Anteil, Konsens-Modal.

auch die neu entwickelten targetbasierten Scorings der Profilauswertung und aus dem Signalentdeckungsansatz betrachtet.

Tabelle 21: Deskriptive Statistiken des ViTEIP-V2 mit den Antwortformaten „Ankreuzen“ und „Zuordnen“ bei unterschiedlichen Scorings

Kennwert	Scoring	„Ankreuzen“ (Stichprobe 2a)	„Zuordnen“ (Stichprobe 2b)	Mittelwertvergleich
M	Target-Übereinstimmung	194.41	202.30	$d=0.34$ ($t=-4.29^*$) ¹
	Target-Treffer	33.16	40.47	$d=0.54$ ($t=-10.22^*$)
	Konsens-Anteil	35.64	44.17	$d=0.72$ ($t=-20.00^*$) ¹
	Konsens-Modal	46.28	58.39	$d=0.81$ ($t=-12.88^*$) ¹
	Profil-Gesamtwert	5.38	5.56	$d=0.02$ ($t=-1.43$)
	Detektionsleistung	0.53	0.66	$d=0.01$ ($t=-3.19$) ¹
	Emotionale Sensitivität	122.22	111.74	$d=-0.65$ ($t=9.88$) ¹
	Emotionale Spezifität	72.18	90.56	$d=0.77$ ($t=-9.63$) ¹
s²	Target-Übereinstimmung	268.39	175.38	
	Target-Treffer	35.04	31.00	
	Konsens-Anteil	5.09	17.71	
	Konsens-Modal	33.21	78.42	
	Profil-Gesamtwert	1.19	0.94	
	Detektionsleistung	0.13	0.09	
	Emotionale Sensitivität	73.80	72.33	
	Emotionale Spezifität	277.45	199.27	
α	Target-Übereinstimmung	0.75	0.58	
	Target-Treffer	0.48	0.46	
	Konsens-Anteil	0.59	0.76	
	Konsens-Modal	0.33	0.73	
	Profil-Gesamtwert	0.31 (0.78)	0.12 (0.53)	
	Detektionsleistung	k.A.	k.A.	
	Emotionale Sensitivität	0.63	0.57	
	Emotionale Spezifität	0.87	0.80	

Anmerkungen:

Für eine detaillierte Beschreibung der Scorings siehe Kapitel 3.3.1 bzw. 4.2

„Ankreuzen“ bzw. „Zuordnen“ = Antwortformate des ViTEIP, Beschreibung siehe Kapitel 1.5.3

M = Mittelwert. s² = Varianz. α = interne Konsistenz Cronbachs α, berechnet über Items; in Klammern angegeben ist die Reliabilität nach Spearman-Brown-Korrektur zur Testverlängerung auf 99 Items bei der Profilauswertung. Mittelwertvergleiche zwischen Stichproben: t = t-Test für unabhängige Stichproben, zweiseitige Testung zwischen je zwei Stichproben; d = Effektstärke.

k.A. = Berechnung der internen Konsistenz nicht möglich.

¹ Levene-Test der Varianzgleichheit auf dem 0.05 Niveau signifikant.

*Auf dem Niveau 0.05 signifikant.

Deskriptive Statistiken und Mittelwertvergleiche. Die Stichprobe 2b, die den ViTEIP-V2b mit dem Antwortformat „Zuordnen“ bearbeitete, erreichte bei den ‚klassischen‘ Scorings höhere Mittelwerte (Tabelle 21) als Stichprobe 2a; die Mittelwerte unterscheiden sich signifikant mit Effektstärken zwischen $d \geq 0.34$ und $d \leq 0.81$. Wurden die neu entwickelten targetbasierten Scorings eingesetzt und die Kennwerte „Profil-Gesamtwert“ sowie „Detektionsleistung“ betrachtet, nivellierten sich die Unterschiede zwischen den Stichproben. Beide Ansätze reduzierten somit den Einfluss, den das unterschiedliche Antwortformat auf die Testleistung ausmacht. Für die getrennte Auswertung, die

zu den Kennwerten „Emotionale Sensitivität“ und „Emotionale Spezifität“ führt, zeigten sich zwischen den Teilstichproben Unterschiede in den Mittelwerten. Die Richtung der Unterschiede wies in die Richtung, dass der erstgenannte Kennwert in der Stichprobe 2a größer ausfiel. Diese Stichprobe bearbeitete das Antwortformat „Ankreuzen“. Stichprobe 2b erreichte höhere Werte in Emotionaler Spezifität, d.h. in den Items, in denen es darum ging, eine korrekte Aussage bezüglich nicht vorhandener Emotionen zu treffen. Für das Antwortformat „Ankreuzen“ ergaben sich für die targetbasierte Scorings größere Varianzen als bei „Zuordnen“.

Für konsensbasierten Scorings zeigten sich bei „Ankreuzen“ mit Werten von $s^2=5.09$ für Konsens-Anteil bzw. $s^2=33.21$ für Konsens-Modal deutlich geringere Varianzen als bei entsprechenden Scorings bei „Zuordnen“ ($s^2=17.71$ Konsens-Anteil, $s^2=78.42$ Konsens-Modal). Die Reliabilitäten lagen bei allen targetbasierten Scorings des Antwortformates „Ankreuzen“ etwas höher als bei „Zuordnen“. Entgegengesetzt verhält es sich bei den Konsensscorings: Hier fielen die Reliabilitäten bei „Zuordnen“ zum Teil deutlich höher aus als bei „Ankreuzen“. Auffällig sind die Unterschiede in der Reliabilität bei Konsens-Modal, mit einer sehr geringen internen Konsistenz von $\alpha=0.33$ bei „Ankreuzen“ und einer höheren internen Konsistenz von $\alpha=0.73$ bei „Zuordnen“.

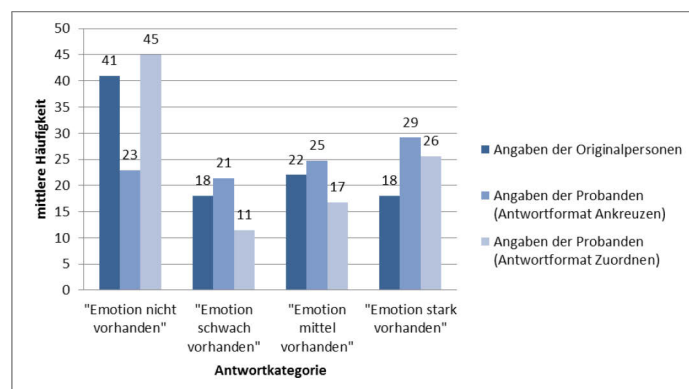


Abbildung 11: Darstellung der mittleren Verteilung der Antwortkategorien im ViTEIP-V2

Um zu prüfen, wie sich die Antwortformate „Ankreuzen“ und „Zuordnen“ auf das generelle Ankreuzverhalten der Probanden auswirkten, wurde in den Teilstichproben eine Auszählung der Häufigkeiten vorgenommen, mit der die unterschiedlichen Antwortkategorien (d.h. „Emotion nicht vorhanden“, „Emotion schwach vorhanden“, usw.) im Durchschnitt gewählt wurden. In *Abbildung 11* ist dargestellt, wie die mittlere Häufigkeit der einzelnen Antwortkategorien ausfällt. Als Vergleichsgröße ist zudem die Anzahl angegeben, wie sich die Antwortkategorien in den Angaben

der Originalpersonen²⁷ verteilen. So finden sich beispielsweise im ViTEIP-V2 insgesamt 41 Items, bei denen die Originalpersonen „Emotion ist nicht vorhanden“ angaben und 22 Items mit der Angabe „Emotion mittel vorhanden“. Es zeigte sich, dass die Probanden bei der Bearbeitung des ViTEIP-V2 mit Antwortformat „Ankreuzen“ (jeweils mittlerer Balken des Diagramms) alle vier Antwortoptionen mit durchschnittlichen Werten zwischen 21 und 29 in etwa gleich häufig wählten. Bei Antwortformat „Zuordnen“ (jeweils rechter Balken) stieg die Anzahl der Häufigkeit, mit der die Probanden die Antwortoption „Emotion nicht vorhanden“ wählten deutlich an (von 23 bei „Ankreuzen“ auf 45 bei „Zuordnen“). Entsprechend seltener wurden die anderen Antwortkategorien gewählt; die größte negative Veränderung war in der Kategorie „Emotion schwach vorhanden“ zu sehen mit durchschnittlich 21 Wahlen bei „Ankreuzen“ und nur durchschnittlich 11 Wahlen bei „Zuordnen“. Mit der Veränderung des Antwortformates ging somit ein verändertes Antwortverhalten der Probanden einher²⁸. Dies kommt bei unterschiedlichen Auswertungsansätzen in unterschiedlicher Intensität zum Tragen.

Prüfung der Unterschiedlichkeit der Korrelationen. Es wurde geprüft, ob sich bei den Antwortformaten „Ankreuzen“ und „Zuordnen“ signifikante Unterschiede in den Korrelationen zwischen ViTEIP-V2 und Kennwerten anderer Verfahren ergeben. Die Prüfung der Unterschiedlichkeit zweier Korrelationskoeffizienten in zwei Stichproben erfolgte nach der Formel von Bortz (2005, S. 222-223); siehe *Formel 13 im Anhang F*. Die Korrelationen lagen fast durchgängig unter dem kritischen Prüfwert von $z_{krit}=\pm 1.96$; die Korrelationen des ViTEIP (bei unterschiedlichen Scorings) zu anderen Verfahren fielen in beiden Stichproben vergleichbar aus (*Anhang I, Tabelle A-32*).

4.4 Vergleich der Präsentationsmodi

Um Auswirkungen des Präsentationsmodus zu prüfen, wurden deskriptive Kennwerte der text- und der videobasierten Forschungsversion (mit je 55 Items) betrachtet. Die Prüfung der Mittelwertunterschiede erfolgte mit *t*-Tests für verbundene Stichproben und der zugrundeliegenden Hypothese, dass es keine Unterschiede in den beiden Testversionen gibt. Es wurde berechnet, ob sich die Korrelationen zu anderen Verfahren bei den beiden Präsentationsmodi unterscheiden. Verwendung fanden die nichtstandardisierten Daten. Aufgrund der geringen Itemanzahl in der Forschungsversion ist eine Auswertung über Profile sowie die Übertragung des Signalentdeckungs-

²⁷ Originalpersonen= Personen, die im Rahmen der Testkonstruktion ihre Situation als Basis für den Test geschildert haben; siehe Kapitel 1.5 zur Entwicklung des ViTEIP.

²⁸ Diese Ergebnisse beziehen sich auf die Ebene der Teilstichprobe; eine Aussage darüber, inwiefern sich das veränderte Antwortformat auf Probandenebene auswirkt, ist aufgrund des gewählten Untersuchungsdesigns nicht möglich. Um darüber Auskunft zu geben, müsste jeder Proband den ViTEIP-V2 mit beiden Antwortformaten bearbeiten, so dass die intraindividuellen Unterschiede betrachtet werden können.

ansatzes, in dem zwischen Null- und Nicht-Null-Items unterschieden wird, nicht sinnvoll; betrachtet werden die Scorings Target-Übereinstimmung, Target-Treffer, Konsens-Anteil und Konsens-Modal.

Itemebene. Die Mittelwerte der text- und der videobasierten Version unterschieden sich auf Ebene der Items zum Teil deutlich voneinander (*Anhang J, Tabellen A-33 bis A-36*). Bei Scoring Target-Übereinstimmung fallen die *t*-Tests für 13 Items signifikant aus, bei Target-Treffer und Konsens-Modal für jeweils 10 und bei Konsens-Anteil sogar für 25 Items. Die Effektgrößen liegen nach der Klassifizierung von Cohen (1992) in der Regel im geringen bis mittleren Bereich. Ausnahmen bilden die Ergebnisse bei Konsens-Anteil, wo für sechs von insgesamt 25 signifikanten *t*-Tests die Effektstärken über $d \geq 0.50$ liegen und auf große Unterschiede zwischen „Text“ und „Video“ hinweisen.

Gesamttestebene. Auf Gesamttestebene fanden sich hingegen zwischen den beiden Versionen „Text“ und „Video“ bei unterschiedlichen Scorings nur sehr kleine Unterschiede in den Kennwerten (*Anhang J, Tabelle A-37*). Folgende Effektstärken zwischen den Versionen wurden berechnet: Target-Übereinstimmung²⁹ $d=0.23$ ($t=3.02^*$); Target-Treffer $d=0.17$ ($t=2.32^*$); Konsens-Anteil $d=0.06$ ($t=0.74$) und Konsens-Modal $d=0.19$ ($t=2.18^*$). Die Effektstärken von maximal $d \leq 0.23$ weisen auf einen geringen Unterschied in den Mittelwerten zugunsten der Textversion bei Auswertung mit Target-Übereinstimmung und Target-Treffer; bei Scoring mit Konsens-Modal erreichten die Probanden etwas bessere Ergebnisse im videobasierten Test. Bei targetbasierten Auswertungen fielen die Varianzen für „Video“ höher aus, bei konsensbasierten Auswertungen hingegen für „Text“. Die Reliabilitäten liegen mit $\alpha \geq 0.63$ im mittleren Bereich; eine Ausnahme bildete die videobasierte Version bei der Auswertung Konsens-Modal mit dem geringeren Wert von $\alpha=0.44$. In Bezug auf die Akzeptanz (*Anhang J, Tabelle A-38*) gab es für beide Präsentationsmodi keine Unterschiede in den Zustimmungswerten.

Prüfung der Unterschiedlichkeit der Korrelationen. Die Zusammenhänge des ViTEIP im Format „Video“ bzw. „Text“ zum TEMINT unterschieden sich nur in drei Korrelationen signifikant, weitere Korrelationen lagen jedoch zum Teil nur knapp unter dem kritischen *z*-Wert (*Tabelle 22*). Wurde der TEMINT mit targetbasierten Ansätzen ausgewertet, fielen die Korrelationen zum ViTEIP (sowohl bei „Text“ als auch „Video“) negativ aus; wurde der TEMINT mit konsensbasierten Scorings ausgewertet fielen diese Zusammenhänge positiv aus. Sowohl für verbale Intelligenz als auch für Merkfähigkeit und Aufmerksamkeit gab es keine signifikanten Unterschiede in der Korrelationshöhe. Es zeigten sich nur wenige signifikante Unterschiede in den Zusammenhängen zu anderen Verfahren oder Kennwerten. Bei Konsens-Anteil korrelierte – anders als bei der textbasierten Version – die Video-Version signifikant mit Verträglichkeit sowie mit zwei EKF-Facetten. In Bezug auf Korrelationen zu

²⁹ *t*= T-Test bei verbundenen Stichproben, $df=112$. d = Effektstärke. *Auf dem Niveau 0.05 signifikant.

demografischen Variablen gab es keine signifikanten Unterschiede zwischen den beiden Präsentationsmodi. Zwischen den beiden Versionen „Text“ und „Video“ lagen die Korrelationen für die verschiedenen Scorings bei $r_{T-U}=0.69^{**}$, $r_{T-T}=0.70^{**}$, $r_{K-A}=0.71^{**}$ sowie $r_{K-M}=0.58^{**}$ (Anhang J, Tabelle A-39).

Tabelle 22: Korrelationen der video- und textbasierten Version des ViTEIP mit anderen Kennwerten sowie Prüfung auf Unterschiedlichkeit der Korrelationen

	Target-Übereinstimmung			Target-Treffer			Konsens-Anteil			Konsens-Modal		
	„Text“	„Video“	z	„Text“	„Video“	z	„Text“	„Video“	z	„Text“	„Video“	z
TEMINT T-Ü	-0.34**	-0.22*	-1.76	-0.30**	-0.14	-2.18	-0.27**	-0.14	-1.91	-0.39**	-0.29**	-1.27
TEMINT T-T	-0.29**	-0.16*	-1.86	-0.23**	-0.08	-2.02	-0.19*	-0.05	-1.97	-0.32**	-0.23**	-1.03
TEMINT K-A	0.06	0.18*	-1.61	0.14	0.24**	-1.35	0.28**	0.38**	-1.46	0.10	0.17*	-0.81
TEMINT K-M	0.02	0.13	-1.45	0.07	0.19*	-1.74	0.17*	0.26**	-1.28	0.00	0.07	-0.79
d2-R_KL	0.08	0.03	0.67	0.07	0.00	0.95	-0.01	0.01	-0.28	-0.02	0.01	-0.34
IST_Merkf.	0.19*	0.17	0.27	0.16	0.11	0.69	0.13	0.05	1.10	0.17	0.11	0.70
IST_Verbal	0.25**	0.36**	-1.55	0.23*	0.32**	-1.28	0.15	0.13	0.28	0.18	0.14	0.47
IST_Figural	0.13	0.09	-3.18	0.13	0.05	1.09	0.03	0.00	0.41	0.09	-0.09	2.08
NEO-FFI_N	-0.17	-0.14	-0.41	-0.13	-0.13	0.00	-0.14	-0.03	-1.52	-0.15	0.01	-1.84
NEO-FFI_E	0.00	-0.07	0.93	0.03	-0.04	0.95	-0.07	-0.06	-0.14	-0.12	-0.14	0.23
NEO-FFI_O	0.03	0.09	-0.80	0.09	0.08	0.14	0.03	0.08	-0.69	0.05	0.07	-0.23
NEO-FFI_V	-0.02	0.03	-0.67	0.06	0.05	0.14	0.03	0.19*	-2.22	0.05	0.17	-1.39
NEO-FFI_G	0.15	0.11	0.54	0.16	0.11	0.68	0.07	0.07	0.00	0.00	0.04	-0.46
EKF_Gesamt	0.14	0.22*	-1.09	0.23*	0.24*	-0.14	0.09	0.20*	-1.54	0.08	0.16	-0.92
EKF_EE	0.11	0.23*	-1.63	0.21*	0.22**	-0.14	0.04	0.25**	-2.94	0.06	0.18	-1.39
EKF_EA	0.02	0.12	-1.34	0.07	0.13	-0.82	0.05	0.20*	-2.09	0.03	0.19	-1.85
EKF_RG	0.13	0.23*	-1.36	0.13	0.21*	-1.10	0.11	0.08	0.42	0.14	0.09	0.58
EKF_EX	0.08	0.03	0.67	0.13	0.05	1.09	0.04	0.01	0.41	-0.01	0.00	-0.11
Alter ¹	0.00	-0.12	1.60	-0.01	-0.07	0.81	0.01	-0.03	0.55	-0.01	-0.07	0.69
Geschlecht ²	0.12	0.04	1.07	0.14	0.10	0.55	0.04	0.07	-0.41	0.15	0.11	0.46
Erfahrungen ³	-0.07	0.05	-1.60	-0.07	0.07	-1.90	-0.02	0.10	-1.66	-0.08	0.04	-1.38
Kontakt ⁴	-0.17	-0.16	-0.14	-0.13	-0.15	0.27	-0.05	-0.07	0.28	-0.08	-0.09	0.11

Anmerkungen:

Korrelation nach Pearson. Einseitige Testung, * $p<0.05$. ** $p<0.01$.

Scoring: T-Ü= Target-Übereinstimmung; T-T= Target-Treffer; K-A= Konsens-Anteil; K-M= Konsens-Modal. Für eine detaillierte Beschreibung der Scorings siehe Kapitel 3.3.1

z= Prüfgröße zur Prüfung der Unterschiedlichkeit der Korrelationen zwischen der video- und der textbasierten Version zu den anderen Kennwerten; kritischer Wert auf dem 5%-Niveau bei zweiseitigem Test $z_{krit}=\pm 1.96$. Zur besseren Übersichtlichkeit wurden signifikante Unterschied **fett** gedruckt.

d2-R= Aufmerksamkeits-Konzentrations-Test d2, revidierte Form (Brickenkamp, Schmidt-Atzert, & Liepmann, 2010), Kennwert Konzentrationsleistung. IST= Intelligenz-Struktur-Test 2000-R (Amthauer, Brocke, Liepmann, & Beauducel, 2001). NEO-FFI= NEO-Fünf-Faktoren Inventar nach Costa und McCrae (Borkenau & Ostendorf, 1993); N= Neurotizismus; E= Extraversion; O= Offenheit für Neues; V= Verträglichkeit; G= Gewissenhaftigkeit. EKF= Emotionale-Kompetenz-Fragebogen (Rindermann, 2007). EKF_EE= EKF-Skala Erkennen eigener Emotionen; EKF_EA= EKF-Skala Erkennen Emotionen anderer; EKF_RG= EKF-Skala Regulation und Kontrolle; EKF_EX= EKF-Skala Emotionale Expressivität.

¹ in Jahren

² Geschlecht (0=männlich, 1=weiblich)

³ Erfahrungen im Pflegebereich in Jahren

⁴ häufiger Kontakt zu älteren Menschen (0=ja, 1=nein)

5 Diskussion

In der vorliegenden Arbeit wurde untersucht, inwiefern sich verschiedene Scorings, veränderte Antwortformate und unterschiedliche Präsentationsmodi auf die Ergebnisse zweier Emotionale-Intelligenz-Tests sowie die Korrelation derselben zu anderen Testverfahren auswirken. *Tabelle 23* bietet eine Übersicht über die Hypothesen und deren Beurteilung (d.h. Bestätigung oder Ablehnung); die differenzierte Diskussion findet sich in den Kapiteln 5.1 bis 5.3. Da sich im Laufe der Analysen einige Fragen bezüglich der Validität der beiden Emotionale-Intelligenz-Tests ergaben, wird diese in Kapitel 5.4 thematisiert. Die Diskussion endet mit Anregungen für zukünftige Studien und einer Auseinandersetzung mit den Limitationen der vorliegenden Untersuchung.

Tabelle 23: Tabellarischer Überblick über die Hypothesen und deren Beurteilung

Forschungsfrage	Nr.	Hypothese	Beurteilung ¹
Auswertungsstrategien			
Vergleich von target- und konsensbasierten Scorings	1a	Korrelation zwischen Scorings signifikant auf dem Niveau $p < 0.01$	Hypothese wird angenommen.
	1b	Korrelation zu anderen Verfahren: Konsens = Target	Hypothese wird angenommen.
Neue Scorings mit biasfreien Kennwerten	1c	Antwortbias bei „Profil-Gesamt“ geringer als bei „Target-Übereinstimmung“	Hypothese wird angenommen.
	1d	Antwortbias bei „Detektionsleistung“ geringer als bei „Target-Übereinstimmung“	Hypothese wird angenommen.
		<i>Post-hoc Ergänzung: Antwortbias bei „Emotionale Sensitivität und Emotionale Spezifität“ geringer als bei „Target-Übereinstimmung“</i>	Hypothese wird angenommen.
	1e	Bestätigung der konvergenten Validität	Hypothese wird (über die Tests und Stichproben hinweg betrachtet) abgelehnt.
Antwortformate	2a	Zentrale Kennwerte: Ankreuzen = Zuordnen	Hypothese wird abgelehnt für „klassische“ Scorings. Hypothese wird angenommen für „neue“ Scorings.
	2b	Korrelationsmuster: Ankreuzen = Zuordnen	Hypothese wird angenommen.
Präsentationsmodus	3a	Korrelation mit verbaler Intelligenz: Text > Video	Hypothese wird abgelehnt.
	3b	Akzeptanz des Verfahrens durch die Probanden: Video > Text	Hypothese wird abgelehnt.
	3c	Korrelation mit Merkfähigkeit: Video > Text	Hypothese wird abgelehnt.
	3d	Korrelation mit Aufmerksamkeit: Video > Text	Hypothese wird abgelehnt.

Anmerkungen:

Nr. = Nummerierung der Hypothesen.

„klassische“ Scorings= Target-Übereinstimmung, Target-Treffer, Konsens-Anteil, Konsens-Modal.

„neue“ Scorings= Profilauswertung sowie Scorings, die aus der Signalentdeckungstheorie abgeleitet wurden (Detektionsleistung, Emotionale Sensitivität und Emotionale Spezifität).

¹ Ausführliche Erläuterung im nachfolgenden Text.

5.1 Auswertungsstrategien

In der Literatur zu Emotionaler Intelligenz werden grundsätzlich drei Ansätze zur Generierung und Beurteilung der „richtigen“ Antwort vorgestellt: das Target-, das Konsens- und das Expertenscoring. Wie in Kapitel 1.2.1 beschrieben, weisen alle drei Herangehensweisen Vor- und Nachteile auf. Inwiefern sich die Wahl des Auswertungsansatzes auf die Ergebnisse der Probanden in einem Emotionale-Intelligenz-Test und auf die Validität auswirkt, bleibt in der Literatur in der Regel unerwähnt; die wenigen Studien³⁰, die Scorings systematisch untersuchen, kommen diesbezüglich zu divergierenden Ergebnissen und unterschiedlichen Interpretationen (Geher, Warner, & Brown, 2001; Mayer, Caruso, & Salovey, 2000). Die Frage, die im Rahmen der vorliegenden Dissertation untersucht wird und aus der die Hypothesen abgeleitet wurden, lautet in allgemeiner Formulierung: Was passiert, wenn ein Emotionale-Intelligenz-Test mit unterschiedlichen Scorings ausgewertet wird? Ausgehend von den Ergebnissen, die sich im Kontext der Beantwortung dieser Frage ergeben, werden Empfehlungen ausgesprochen, welche Scorings sich für Tests zur Erfassung von Emotionaler Intelligenz anbieten bzw. aus Sicht der Autorin zu präferieren sind.

5.1.1 Vier target- und konsensbasierte Scorings im Vergleich

Die *Hypothesen 1a und 1b* sind in Form von Nullhypothesen formuliert mit der Annahme, dass die Wahl eines anderen Auswertungsansatzes keinen bedeutsamen Einfluss auf die Testergebnisse hat. Wie unter 2.1 beschrieben, bezieht sich diese Annahme darauf, dass eine geänderte Auswertungsmethodik zwar die absolute Höhe der Testkennwerte innerhalb einer Stichprobe verändern kann, nicht jedoch die relative Position eines Probanden im Vergleich zu den anderen Probanden der Stichprobe beeinflussen sollte. Andernfalls müsste die Auswertungsmethode zu der Kategorie der systematischen Methodeneffekte (Cote & Buckley, 1987; Spector, 2006) gerechnet werden. Systematische Methodeneffekte bieten eine alternative Erklärung für beobachtete Zusammenhänge und erschweren eine Aussage über das zu erfassende Konstrukt (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Die gleichberechtigte Darstellung der drei Scorings in der Literatur zu Emotionaler Intelligenz ohne Hinweise auf mögliche Unterschiede zwischen denselben, weist in Richtung der beschriebenen Annahme – ob berechtigt, soll mit der vorliegenden Arbeit untersucht werden. Die Hypothesen wurden somit aus einem testtheoretisch-methodischen Blickwinkel formuliert; inhaltliche Aspekte, wie beispielsweise die grundlegende Logik, die hinter verschiedenen Scorings steht oder Ergebnisse weiterer Studien oder mit anderen Konstrukten, finden in der Diskussion ausreichend Berücksichtigung.

³⁰ An dieser Stelle soll auf die File-drawer-Problematik verwiesen werden, da nur veröffentlichte Studien gesichtet werden können. Inwiefern entsprechende Untersuchungen ohne eine Veröffentlichung erfolgten, kann nicht beurteilt werden.

Zur Überprüfung der Hypothesen wurden die Emotionale-Intelligenz-Tests ViTEIP-V2 und TEMINT zunächst mit vier verschiedenen Scorings ausgewertet und die Ergebnisse miteinander verglichen. Zum Einsatz kamen die Scorings Target-Übereinstimmung und Target-Treffer, bei denen sich die Testwerte daraus generieren, dass die Angabe des Probanden mit der Angabe der Originalperson – also dem Target – verglichen und ein Übereinstimmungs- bzw. Trefferwert berechnet wurde. Bei den konsensbasierten Scorings Konsens-Anteil und Konsens-Modal wurde der Gesamtwert aus der Übereinstimmung mit der Mehrheitsmeinung der Stichprobe generiert. Eine genaue Beschreibung dieser Scorings findet sich in Kapitel 3.3.1. Die Berechnungen wurden an den Ergebnissen der studentischen Stichprobe 1 und der Teilstichprobe 2a (Altenhilfeschüler) durchgeführt, die beide den ViTEIP-V2 mit demselben Antwortformat „Ankreuzen“ bearbeiteten.

Deskriptive Statistiken und Faktorstruktur. Vor der Beantwortung der Hypothesen soll zunächst ein Blick auf allgemeine Kennwerte der Tests sowie die faktorielle Struktur bei unterschiedlichen Scorings geworfen werden. Es zeigte sich, dass die Varianz bei ViTEIP-V2 und TEMINT bei einer Auswertung mit Target-Übereinstimmung jeweils am größten ausfiel (*Tabelle 7*). Aufgrund der geringen Varianzen bei konsensbasierten Auswertungen ist davon auszugehen, dass die Messwerte hier nur eine geringe Dispersion um den Mittelwert aufweisen. Eine höhere Varianz ist generell wünschenswert und ein Indikator dafür, dass ein Test die Unterschiede in der Leistungsfähigkeit der Probanden abbildet und damit zwischen Personen differenzieren kann. Varianzeinschränkungen hingegen können dazu führen, dass vorhandene Zusammenhänge zwischen Variablen bzw. Personen nicht identifiziert werden. Auch in den Untersuchungen von Geher et al. (Geher, Warner, & Brown, 2001) für das Instrument EARS und bei Mayer et al. (Mayer, Caruso, & Salovey, 2000) für den MSCEIT fiel die Varianz bei targetbasierter Auswertung größer aus als bei Konsens, was an sich für den Einsatz targetbasierter Scorings spricht.

Sowohl für den ViTEIP-V2 als auch TEMINT wurden mit Target-Übereinstimmung die höchsten Reliabilitäten erreicht. Hier kann die höhere Varianz der Items bei targetbasierter Auswertung einen Einfluss haben, da eine höhere Streuung in der Regel auch zu höheren Reliabilitäten führt. Die Ergebnisse reihen sich ein in die Serie der Arbeiten, in denen sich bei verschiedenen Auswertungsansätzen Unterschiede in den Reliabilitäten zeigen (Geher, Warner, & Brown, 2001; Mayer, Caruso, & Salovey, 2000; Mayer & Geher, 1996; MacCann, Roberts, Matthews, & Zeidner, 2004). In der vorliegenden Arbeit fielen die internen Konsistenzen bei Target deutlich größer aus als bei Konsens; dies beschreiben auch Davies et al. (Davies, Stankov, & Roberts, 1998). In den anderen zitierten Studien war die höhere Reliabilität überwiegend bei einer konsensbasierten Auswertung zu finden. Insgesamt wurde deutlich, dass das mittels interner Konsistenz dargestellte Ausmaß, in dem die Fragen eines Tests miteinander in Beziehung stehen, bei unterschiedlichen Scorings variiert.

Der ViTEIP wurde mit der Annahme konzipiert, mit der Fähigkeit „Erkennen von Emotionen Anderer anhand situativer Auslöser“ einen homogenen Merkmalsbereich des Konstrukts Emotionale Intelligenz zu erfassen. Da die Kenntnis der internen Konsistenz zwar Aussagen über die Messgenauigkeit eines Tests ermöglichen, jedoch nicht über die Homogenität (Cortina, 1993), wurden exploratorische Faktorenanalysen berechnet. Aufgrund der Tatsache, dass die Daten auf Itemebene nicht für faktorenanalytische Berechnungen geeignet waren³¹, ist die Annahme der Homogenität in Frage zu stellen. Eine mögliche Erklärung für die geringen Interkorrelationen zwischen den Items findet sich bei Austin (2010). Sie regte an, nicht ungeprüft vorauszusetzen, dass alle Emotionen von Probanden gleichermaßen verarbeitet werden können; es bräuchte eigentlich emotionshomogene Tests. Aus ökonomischer Sicht ist dieser Vorschlag jedoch kritisch zu sehen, da der Aufwand für Testentwicklung und später für die Testadministration als enorm einzuschätzen ist. Auch ist der Nutzen einer emotionsspezifischen Rückmeldung im Rahmen einer diagnostischen Testung im Sinne von „Frau X kann Ärger, Trauer und Freude überdurchschnittlich gut, Stolz, Scham und Überraschung unterdurchschnittlich gut erkennen“ fraglich.

Als nächster Analyseschritt wurden die Items zu inhaltlichen Clustern zusammengefasst, die der Zugehörigkeit zu der von den Originalpersonen geschilderten Situation entsprechen. Auch auf dieser Aggregationsebene ist das Vorliegen eines homogenen Konstrukts Emotionale Intelligenz zu bezweifeln, da es zwischen den verschiedenen Auswertungsmethoden keine Übereinstimmung darüber gab, welche Situationen auf den jeweiligen Komponenten/ Faktoren luden (*Tabellen A-6 bis A-23, Anhang E*); dies gilt auch für den TEMINT. Es zeigte sich zwar in den Hauptkomponentenanalysen jeweils eine erste starke Komponente, die zwischen 17 und 26% der Varianz aufklären konnte. Für ein homogenes Konstrukt erscheint diese Varianzaufklärung aus Sicht der Autorin jedoch zu gering. Wie in Kapitel 1.4 beschrieben, fielen auch in anderen Untersuchungen zur faktoriellen Struktur von Emotionale-Intelligenz-Tests die Ergebnisse heterogen aus. So kamen zum Beispiel für den häufig eingesetzten MSCEIT verschiedene Autoren zu unterschiedlichen Faktorstrukturen – mal wird eine drei-, mal eine vierfaktorielle Lösung empfohlen, mal eine Generalfaktorstuktur. Für den STEM und STEU waren die Ergebnisse zur faktoriellen Zuordnung von Items ebenso uneinheitlich (Ferguson & Austin, 2011). Sharma et al. (Sharma, Gangopadhyay, Austin, & Mandal, 2013) verstehen Emotionale Intelligenz grundsätzlich nicht als einheitliches Konstrukt und schlussfolgern, dass daher der Einsatz von Situational Judgment Tests für die Erfassung am geeignetsten sei.

³¹ Als Indikator für die Dateneignung wurde der Kaiser-Meyer-Olkin-Koeffizient (KMO) betrachtet: Ist der Koeffizient gering, teilen die Items nur wenig gemeinsame Varianz und die Daten sind nicht für Faktorenanalysen geeignet.

Insgesamt zeigten sich im Rahmen der Faktorenanalysen sowohl für den ViTEIP-V2 und TEMINT als auch für in anderen Studien verwendete Emotionale-Intelligenz-Tests, dass entweder das Konstrukt Emotionale Intelligenz eine hohe Heterogenität besitzt und/oder existierende Verfahren zur Messung derselben noch keine ausreichende psychometrische Qualität erreicht haben, um das Konstrukt entsprechend abzubilden. Vor allem aufgrund der Heterogenität der Ergebnisse (Divergenzen in der Zuordnung der Situationen auf die Komponenten) in den verschiedenen Stichproben und bei unterschiedlichen Scorings wurde deutlich, dass möglicherweise weitere Faktoren identifiziert werden müssen, die bei der Identifikation von Emotionen anhand situativer Hinweise eine Rolle spielen. Ein möglicher Ansatz besteht darin zu untersuchen, inwiefern beispielsweise die Unterscheidung in positive/angenehme und negative/unangenehme Emotionen erforderlich ist, um Ergebnisse zu erklären. Für die Emotionserkennung in Gesichtern gibt es hierzu divergierende Ergebnisse (Schlegel, 2013; Suzuki, Hoshino, & Shigemasu, 2010). Mit den in der vorliegenden Arbeit neu entwickelten Kennwerten Emotionale Sensitivität und Emotionale Spezifität wurde mit der getrennten Auswertung nach Null- und Nicht-Null-Items³² erfolgreich ein relevantes Unterscheidungsmerkmal identifiziert (siehe Kapitel 5.1.2.3).

Korrelationsanalysen. Zunächst werden die Korrelationen zwischen den Scorings innerhalb desselben Tests betrachtet, daran anschließend die Korrelationsmuster zwischen den Verfahren. Es folgt eine übergreifende Diskussion.

Die *Hypothese 1a* „Die unterschiedlichen Scorings sind miteinander vergleichbar, was sich in einer hohen positiven Korrelation ($**p < 0.01$) äußert“ wird angenommen. Die signifikant positiven Korrelationen ($r \geq 0.48^{**}$, Tabelle 8; Ausnahme: Stichprobe 2a ViTEIP-V2 mit einer geringeren Korrelation) zwischen den Scorings innerhalb desselben Tests zeigen, dass die Testergebnisse bei unterschiedlichen Auswertungen einen gemeinsamen Varianzanteil besitzen und grundsätzlich vergleichbar sind. Leistungsunterschiede werden somit mit verschiedenen Scorings in entsprechender Weise deutlich. Wenig überraschend fielen dabei die Korrelationen innerhalb der beiden konsensbasierten bzw. der beiden targetbasierten Scorings am höchsten aus. Diesen liegt jeweils dasselbe Grundprinzip zugrunde und unterscheidet sich nur darin, ob Abstufungen in der Bewertung vorgenommen werden oder ob es wie bei Target-Treffer bzw. Konsens-Modal nur eine einzige richtige Antwort gibt. Stichprobe 2a bildete eine Ausnahme in den Ergebnissen, indem die Korrelationen zwischen den verschiedenen Scorings des ViTEIP-V2 deutlich geringer ausfallen ($r \leq 0.35^{**}$). Hier bleiben große Varianzanteile unerklärt. Um sicherzustellen, dass diese Ergebnisse nicht auf die Datenqualität (z.B. auf mögliche falsche Codierungen oder Fehleingaben) oder die

³² Null-Items= Items, bei denen die Antwort der Originalperson „Emotion nicht vorhanden“ lautet. Nicht-Null-Items= Items, bei denen die Antwort der Originalperson „Emotion schwach/mittel/stark vorhanden“ lautet.

Stichprobenzusammensetzung (z.B. Ausreißer, anderes Geschlechterverhältnis innerhalb der Stichprobe etc.) zurückzuführen sind, wurden entsprechende Kontrollen erneut durchgeführt. Diese Aspekte konnten als Ursache für die vergleichsweise geringe Korrelation ausgeschlossen werden.

Die hohen Korrelationen zwischen den Scorings verweisen darauf, dass sich trotz der zuvor beschriebenen Unterschiedlichkeiten in Bezug auf Varianz und Reliabilität die Rangreihe der Ergebnisse der Probanden vergleichbar darstellen und gemeinsame Varianzanteile existieren. Wie auch bei Mayer, Caruso und Salovey (2000) können die hohen Korrelationen zwischen target- und konsensbasierten Auswertungen im Sinne von Übereinstimmungen in den Ergebnissen, die von der Auswertungsprozedur unabhängig sind, interpretiert werden. Die Höhe, wie sie beispielsweise Mayer et al. (Mayer J. D., Salovey, Caruso, & Sitarenios, 2003) für die Korrelation von Expert- und Konsensscoring ($r \geq 0.90$) aufführen, wurden jedoch nur vereinzelt beim TEMINT erreicht.

Im direkten Vergleich der Korrelationshöhen zu anderen Verfahren bei unterschiedlichen Auswertungen zeigten sich für beide Tests nur einige wenige signifikante Unterschiede (*Tabellen 9 und 10*) in Stichprobe 1. Die *Hypothese 1b* „Unterschiedliche Auswertungsmethoden haben keine Auswirkungen auf die Korrelationsmuster in Bezug auf die verschiedenen Validitätsaspekte“ wird damit bestätigt. Zu beachten ist jedoch, dass die beobachteten Korrelationen nur teilweise mit den erwarteten Korrelationsmustern übereinstimmen und somit grundsätzliche Fragen bezüglich der Validität der eingesetzten Emotionale-Intelligenz-Tests aufwerfen; dies wird in Kapitel 5.4 diskutiert.

Ein besonders auffälliges – und sowohl unerwartetes als auch der Erwartung widersprechendes – Ergebnis stellte der Zusammenhang zwischen den beiden Emotionale-Intelligenz-Tests ViTEIP-V2 und TEMINT dar. Die Annahme, dass diese Verfahren im Sinne der konvergenten Validität positiv-signifikant miteinander korrelieren, bestätigte sich nur dann, wenn beide Tests mit konsensbasierten Scorings ausgewertet wurden. Hier waren die Korrelationen im moderaten positiven Bereich um $r=0.30^{**}$. Diese Korrelation entspricht der Höhe, wie sie auch bei Austin (2010) für Korrelationen zwischen STEM und MSCEIT, oder bei Blickle et al. (Blickle G., Momm, Liu, Witzki, & Steinmayr, 2011) zwischen TEMINT und MSCEIT berichtet werden. Ein anderes Bild zeigte sich, wenn ViTEIP-V2 und TEMINT mit targetbasierten Scorings ausgewertet wurden: hier lagen die Korrelationen im Bereich um $r=-0.20^{**}$. Diese Korrelation war zum einen unerwartet, da sie ein negatives Vorzeichen aufweist. Zum anderen ist es insgesamt beachtenswert, dass die Anwendung eines anderen Scorings zu solchen Veränderungen in der Korrelation zwischen Verfahren führt, die dasselbe Konstrukt erfassen sollen. Da diese Änderung der korrelativen Zusammenhänge in dieser Form jedoch nur die beiden Emotionale-Intelligenz-Tests untereinander betraf (*Tabelle 8*) und

überwiegend nicht die Korrelationen der Emotionale-Intelligenz-Tests zu anderen Verfahren (Tabelle 9), wird wie ob beschrieben *Hypothese 1b* dennoch angenommen.

Das unerwartete Korrelationsmuster zwischen den Emotionale-Intelligenz-Tests bei targetbasierten Scorings konnte auf einen Antwortbias auf Seiten der Probanden zurückgeführt werden. Der Antwortbias äußerte sich dergestalt, dass Probanden bei beiden Tests bevorzugt bestimmte Skalenbereiche für ihre Antworten wählten, diese Präferenz jedoch in den beiden Tests zu unterschiedlichen Ergebnissen führte. Dies bedeutet, dass eine Person, die bevorzugt im unteren Skalenbereich „Emotion nicht vorhanden“ antwortete, im ViTEIP-V2 aufgrund der hohen Anzahl an Items, in denen diese Antwort die richtige darstellt, einen hohen Wert bei Target-Übereinstimmung oder Target-Treffer erhielt. Dieselbe Person erhielt mit ihrer Präferenz für diesen Skalenbereich im TEMINT tendenziell schlechtere Testkennwerte, da hier nur wenige Angaben der Originalperson „Emotion nicht vorhanden“ lauten. Für den oberen Skalenbereich liegt die „Bevorzugung“ hingegen beim TEMINT und die „Benachteiligung“ beim ViTEIP-V2. Probanden mit einem ausgeprägten Antwortbias erreichten daher in dem einen Emotionale-Intelligenz-Test gute, in dem anderen Test hingegen schlechtere Ergebnisse. Dies kann auf Stichprobenebene zu der negativen Korrelation zwischen ViTEIP-V2 und TEMINT führen. Im Rahmen der vorliegenden Arbeit wurden neue Scorings entwickelt. Wie im nachfolgenden Kapitel 5.1.2 ausführlich dargestellt wird, waren sowohl die Profilauswertung als auch Scorings aus der Übertragung des Grundgedankens der Signalentdeckungstheorie geeignet, den Einfluss der Antwortpräferenzen zu reduzieren.

Für konsensbasierte Auswertungen ist das Vorliegen eines Antwortbias weniger relevant, da sich die Gewichtung der einzelnen Antworten aus der Verteilung der Antworten der Gesamtstichprobe ergeben. Ein bestimmtes Antwortverhalten kann zwar für einen Probanden im Einzelfall von Nachteil oder Vorteil sein, da er von der Mehrheitsmeinung der Stichprobe abweicht bzw. mit ihr übereinstimmt. Dies ist jedoch Teil der Auswertungsmethodik und keine grundlegende Benachteiligung oder Bevorzugung, wie es bei der targetbasierten Auswertung durch die Festlegung der Korrektheit der Antworten durch die Angabe der Originalperson der Fall ist.

Fazit: Target- und konsensbasierte Scorings. Wie sind diese Ergebnisse insgesamt einzuordnen? Anhand der überwiegend moderaten bis hohen Korrelationen innerhalb der Emotionale-Intelligenz-Tests bei verschiedenen target- und konsensbasierten Scorings sowie der nur geringen Anzahl an signifikanten Korrelationsunterschieden zu anderen Verfahren ist in der vorliegenden Arbeit davon auszugehen, dass es zwischen den Scorings deutliche Überschneidungsbereiche gibt. Das Ergebnis kann in der Hinsicht interpretiert werden, dass die Anwendung eines anderen Scorings, trotz der Unterschiede in Varianz und Reliabilität, zu einer vergleichbaren Rangreihe der Probanden

innerhalb ihrer Stichprobe führt und Leistungsunterschiede erhalten bleiben. Auch die Zusammenhänge zu anderen Verfahren stellten sich in ähnlicher Form dar. Es soll an dieser Stelle jedoch auf die insgesamt geringe Anzahl an signifikanten Korrelationen verwiesen werden, die sich im Rahmen der Validierung ergeben haben. Dadurch können überwiegend Aussagen darüber getroffen werden, was ViTEIP-V2 und TEMINT unter Anwendung verschiedener Scorings *nicht* messen. In welcher Form sich Scoring auswirken, wenn Validitätsbelege klarer ausfallen, bleibt offen.

Aus der eher formalen, testtheoretischen Argumentation heraus ist es wünschenswert, dass sich Testergebnisse bei Anwendung eines anderen Scorings nicht wesentlich ändern. Diese Betrachtungsweise ist nicht allein ausschlaggebend, sondern es sollten in jedem Fall auch inhaltliche Aspekte berücksichtigt werden. So kritisierte auch schon Wilhelm (2005), dass alleine das Vorliegen einer hohen Konvergenz zwischen experten- und konsensbasierten Messungen nicht ausreiche, um diese Scorings für den MSCEIT zu empfehlen. Vielmehr brauche es eine rationale Begründung.

Aufgrund der grundlegenden Unterschiede in der Logik von target- und konsensbasierten Scorings (siehe Kapitel 1.2.1) ist eine hohe Übereinstimmung eher für leichte Items zu erwarten. Wird die Höhe der Korrelation betrachtet, traf dies vor allem für den TEMINT zu, bei dem die Korrelationen im Vergleich zum ViTEIP-V2 noch einmal deutlich höher ausfielen. Im ViTEIP-V2 gab es häufig eine Abweichung zwischen der Angabe der Originalperson und der Antwort, die aus dem Konsensscoring generiert wurde (siehe *Tabellen A-3 und A-4 im Anhang C*). Barchard et al. (Barchard, Hensley, & Anderson, 2013) bestätigten mit ihrer Untersuchung, dass die Lösung, die mittels Konsensscoring identifiziert wurde, bei einfachen und mittelschweren Items eine hohe Korrelation zur objektiv richtigen Antwort besaß. Auch Austin und Saklofske (2006) merkten an, dass konsensbasierte Verfahren eher auf einfache als auf schwierige emotionale Probleme anzuwenden sind. So könnten beispielsweise Gesichtsausdrücke über den Gruppenkonsens ermittelt werden, bei sozialen Situationen, in denen eine überdurchschnittliche Leistung erbracht werden müsse, versage diese Vorgehen (Austin & Saklofske, 2006, S. S. 120). Da es sich vor allem beim ViTEIP-V2 um komplexe Anforderungen der Emotionserkennung aus beschriebenen Situationen handelt, und somit nicht sichergestellt ist, ob mittels Konsens tatsächlich die richtige Antwort identifiziert werden kann, sollte aus Sicht der Autorin auf targetbasierte Auswertungen zurückgegriffen werden.

Die Grundannahme von Konsens, dass Antworten der Probanden auf eine Antwortkategorie konvergieren (Legree, Pstotka, Tremble, & Bourne, 2005) und Konsens somit geeignet ist für Bereiche, in denen keine explizite Expertise und keine eindeutig definierten richtigen Antworten existieren, bestätigte sich in den vorliegenden Daten nur zum Teil. Im TEMINT zeigte sich bei den meisten Items eine hohe Konvergenz auf eine Antwortoption; beim ViTEIP-V2 war dies nur bei

einigen Items der Fall. Bei Target und Konsens waren zum Teil deutliche Abweichungen zu finden, was über die jeweilige Auswertungsmethode als „richtige“ Antwort definiert wird (*Anhang C, Tabelle A-3*). Dass eine Abweichung darüber, welche Antwort als die richtige gewertet wird, nicht ausschließlich kritisch zu sehen ist, darauf verweisen Bergman et al. (Bergman, Drasgow, Donovan, Henning, & Juraska, 2006). In ihrer Publikation verglichen die Autoren verschiedene Scorings eines Situational Judgment Tests zu Führungsfähigkeiten und zeigten auf, welche Antwortoptionen bei welchem Scoring als „korrekt“ bewertet werden. In diesem Zusammenhang erscheint es plausibel, je nach Schwerpunkt der Auswertung (z.B. wird ein bestimmtes Führungsmodell als theoretischer Rahmen zugrunde gelegt? Bestimmen Experten oder Novizen den Auswertungsschlüssel?) zu unterschiedlichen Ergebnissen zu kommen. Für ein Konstrukt wie Emotionale Intelligenz, für das in Anspruch genommen wird, ein Teil eines generellen Intelligenzmodells zu sein, scheint diese „Flexibilität“ bezüglich der korrekten Antwort unangemessen. Im Rahmen der Intelligenzdiagnostik sollte gerade auch bei neuen Intelligenzfacetten nicht von der Forderung nach der Existenz einer objektiv richtigen Antwort abgewichen werden. Wenn unterschiedliche Scorings zu unterschiedlichen Einschätzungen darüber kommen, welche Antwort als korrekt gewertet wird, ist dies aus Sicht der Autorin für eine Einsatz im diagnostischen Bereich kritisch.

Barchard und Russell (2006) weisen darauf hin, dass je nach Zusammensetzung der Referenzstichprobe, anhand derer die Anteilswerte für das Konsensscoring berechnet werden, sich unterschiedliche Auswertungsschlüssel ergeben können. Diese grundlegende Kritik am Konsensscoring wurde in der vorliegenden Arbeit nicht untersucht, bleibt aber ein nach wie vor relevantes Argument gegen den Einsatz von konsensbasierten Scorings. Palmer et al. (Palmer, Gignac, Manocha, & Stough, 2005) konnten zwar zeigen, dass sich für den MSCEIT eine hohe Konvergenz zwischen einer amerikanischen und einer australischen Referenzstichprobe ergab. Da es sich bei Palmer et al. (2005) um zwei Stichproben handelt, die aus vergleichsweise ähnlichen, westlichen Kulturen stammen, kann nicht generell gefolgert werden, der Stichprobeneinfluss sei zu vernachlässigen. Kritiker könnten anmerken, dass das Targetscoring genauso anfällig auf kulturelle Normen sei, wie das Konsensscoring. Dies trifft insofern zu, als dass kulturelle Hintergründe sich sowohl in der Angabe der Originalperson als auch der Probandenantwort wiederfinden können. Gelingt es einem Probanden die kulturspezifischen Bewertungen der Originalperson zu identifizieren, kann durch die Berechnung von Übereinstimmungs- oder Trefferwerten mit dem Target (der Originalperson) eine hohe Testleistung erreicht werden. Die emotional intelligente Leistung eines Probanden besteht also darin, sich trotz unterschiedlicher kultureller Hintergründe in die Lage des Anderen zu versetzen. Bei konsensbasierten Scorings dominiert der kulturelle Hintergrund der sich aus der Zusammensetzung der Referenzstichprobe ergibt.

Da eine targetbasierte Vorgehensweise davon ausgeht, dass nur die betroffenen Personen (Originalpersonen) über die in der Situation/am Ende der Situation vorhandenen Emotionen Auskunft geben können, gilt es die Rahmenbedingungen so zu gestalten, dass die Angaben der Originalperson so valide wie möglich sind. Es wird angeregt Faktoren zu identifizieren, die einen Einfluss auf die Qualität der Angaben der Originalperson haben, wie bspw. Persönlichkeitsfaktoren, kognitive Fähigkeiten, aber auch Faktoren in der Situation, die sich auf das Erinnern auswirken. Erkenntnisse aus der Forschung zu Erinnerungseffekten und zur Qualität von Augenzeugenberichten könnten bei der Planung zukünftiger Untersuchungen berücksichtigt werden. Vor allem die zahlreichen und vielfältigen Untersuchungen von Forgas und Kollegen liefern wertvolle Informationen, welche Aspekte einen Einfluss auf die Erinnerung und den Abruf der Inhalte haben (z.B. Forgas, 2007; Forgas & East, 2008; Forgas, Laham, & Vargas, 2005).

Aus Sicht der Autorin ist es wesentlich, alternative Ansätze zum Konsensscoring zu generieren sowie das Targetscoring weiter zu untersuchen. Aus der geringen Anzahl an Befunden ist nicht nachzuvollziehen, warum die Targetauswertung im Rahmen von Emotionaler Intelligenz so schnell an Bedeutung verloren und sich das Konsens- bzw. Expertenscoring durchgesetzt haben. Die vorliegende Arbeit kann als Anregung und Hinweis an Testautoren verstanden werden, sich intensiv damit auseinanderzusetzen, welches Scoring in einem Test eingesetzt werden soll. Die tatsächlichen Unterschiede bezüglich der Korrelationsmuster fallen zwar insgesamt gering aus, dennoch können target- und konsensbasierte Auswertungsansätze nicht von vorneherein als vergleichbar angesehen werden. Es sei an dieser Stelle noch einmal ausdrücklich auf den Unterschied in der Korrelation zwischen ViTEIP-V2 und TEMINT verwiesen, der sich bei Anwendung von Target bzw. Konsens ergibt. Dieser Unterschied spricht dafür, dass es bei den Scorings Anteile gibt, die in der Anwendung unterschiedlich zum Tragen kommen. Es gelingt in der vorliegenden Arbeit zu zeigen, wodurch diese Korrelationen zustande kommen und wie mit neuen targetbasierten Scorings die tatsächlichen (positiven) Zusammenhänge zwischen den beiden Emotionale-Intelligenz-Tests sichtbar werden.

5.1.2 Zwei neue Scorings mit biasfreien Kennwerten

Bezüglich des Zusammenhangs zwischen den beiden Emotionale-Intelligenz-Tests wurde erwartet, dass ViTEIP-V2 und TEMINT dasselbe Konstrukt erfassen und sich dies in einer positiven Korrelation der Gesamttestergebnisse zeigt. Entgegen dieser Erwartung fielen die Korrelationen zwischen den Tests bei targetbasierten Auswertungen in allen drei Stichproben negativ aus. Als mögliche Ursache für diese negative Korrelation konnte ein Antwortbias identifiziert werden, der auf Seiten der Probanden vorliegt.

Der Antwortbias äußerte sich dergestalt, dass Probanden ihre Antworten in ViTEIP-V2 und TEMINT in einem bevorzugten Skalenbereich wählen. So wählt ein Proband A beispielsweise in ViTEIP-V2 und TEMINT vorzugsweise den unteren Skalenbereich für seine Antworten, ein anderer Proband B entscheidet sich in beiden Tests häufig für Antworten im oberen Skalenbereich. Für beide Probanden wirkt sich ihr Antwortverhalten negativ aus, da in ViTEIP-V2 und TEMINT jeweils unterschiedliche Enden der Antwortskala mit hohen Übereinstimmungs- oder Trefferwerten „belohnt“ werden. Nur Probanden ohne einen präferierten Skalenbereich – d.h. mit einem differenzierten Ankreuzverhalten – haben die Chance, in Abhängigkeit von ihrer Leistungsfähigkeit, in beiden Tests bei targetbasierter Auswertung hohe Gesamtwerte zu erhalten. Der Antwortbias ist dabei nicht gleichzusetzen mit einem Antwortverhalten, das sich aus Sozialer Erwünschtheit oder im Sinne von „Faking“ ergibt, da der ViTEIP-V2 und TEMINT aufgrund des Leistungstestcharakters grundsätzlich so angelegt sind, dass mehr oder weniger bewusste Verfälschungen keinen Vorteil erbringen und maximal Verfälschungen in Richtung schlechterer Leistungsfähigkeit möglich sind. Daher ist davon auszugehen, dass es sich mit dem Antwortbias tatsächlich um eine Art „Disposition“ oder – um an die Begrifflichkeiten der Signalentdeckungstheorie anzuknüpfen – Reaktionsneigung handelt. Aufgrund der möglichen Stabilität des Antwortbias und der Abhängigkeit von anderen demographischen Variablen wie Geschlecht (vgl. Weijters, Geuens, & Schillewaert, 2010), erscheint es dringend erforderlich Scorings zu entwickeln, die die Auswirkungen eines Bias auf Ergebnisse in Emotionale-Intelligenz-Tests reduzieren.

Mit der Profilauswertung und der Übertragung der Signalentdeckungstheorie auf Emotionale-Intelligenz-Tests wurden neue targetbasierte Scorings entwickelt, die gegenüber vorliegenden Antworttendenzen seitens der Probanden deutlich weniger anfällig sind. Vor allem vor dem Hintergrund, dass die bisher bekannten Scorings alle Vor- und Nachteile aufweisen (siehe Kapitel 1.2.1) und gleichzeitig die Suche nach adäquaten Messinstrumenten für Emotionale Intelligenz vorangetrieben wird, liefert die vorliegende Arbeit hier einen entscheidenden Beitrag. Dass im Rahmen der durchgeführten Untersuchungen bezüglich der neuen Scorings noch nicht alle Fragen geklärt werden konnten und weitere Forschung erforderlich ist, ist Bestandteil jeder innovativen Idee.

5.1.2.1 Profilauswertung

Ausgangspunkt für die Profilauswertung waren die identifizierten individuellen Antworttendenzen: Einige Probanden kreuzten tendenziell höhere Werte an und überschätzen somit vorhandene Emotionen, andere Probanden zeigten ein Ankreuzverhalten, das die Emotionsintensität eher unterschätzt. Thomson (2009) prüfte in ihrer Arbeit den Ansatz, Antworttendenzen im Nachhinein

für jeden Probanden individuell zu korrigieren. Durch diese nachträgliche Korrektur konnte der Antwortbias nicht ausreichend reduziert werden; auch erwies sich dieses Vorgehen als wenig praktikabel in der Anwendung. Bei der Profilauswertung hingegen werden keine Korrekturen von Werten vorgenommen, sondern die Testleistung über den Vergleich der Antwortmuster des Probanden und der Originalperson berechnet. Vorteil dieses Ansatzes ist, dass nicht nur der Einfluss möglicher Antworttendenzen auf Probandenseite verringert wird, sondern auch potenzielle Antworttendenzen seitens der Originalperson, die es Probanden gegebenenfalls erschweren die richtige Antwort zu identifizieren. Es ist anzunehmen, dass sich auch die Originalpersonen in ihren Wahrnehmungen und Angaben der Emotionseinschätzungen unterscheiden können und beispielsweise unterschiedliche Maßstäbe an die Intensitätsskala angelegt werden (ab wann wird eine Emotion als „stark vorhanden“ oder „sehr stark vorhanden“ empfunden?). Je nach Wahrnehmung und Einschätzung der Originalperson können auch hier innerhalb der Situationen Antworttendenzen in Richtung stärkere oder schwächere Ausprägungen der Emotionen vorliegen. Mit der Profilauswertung werden auch diese Biases reduziert, da probanden- und targetseitig nicht die absolute Ausprägungshöhe der verschiedenen Emotionen einer Situation berücksichtigt werden, sondern die Rangreihe der Intensitäten. Als Maß für die Übereinstimmung der Antwortprofile von Originalperson und Probanden wurden Rangkorrelationen nach Spearman berechnet (siehe 2.1.1.1). Ein hoher positiver Rangkorrelationswert bedeutet, dass die Rangreihen der Emotionsintensitäten innerhalb der Situation gut mit dem Antwortprofil der Originalperson übereinstimmt. Anders als bei Target-Übereinstimmung oder Target-Treffer führt eine Antworttendenz (z.B. überwiegendes Ankreuzen geringer Emotionsintensitäten) somit nicht grundsätzlich zu besseren oder schlechteren Ergebnissen.

Bei Anwendung der Profilauswertung reduzierte sich für ViTEIP-V2 und TEMINT der Einfluss des Antwortverhaltens der Probanden. Dies zeigte sich in einer mehrheitlich signifikant geringeren Korrelationen zwischen den Kennwerten zum Antwortverhalten und den Gesamttestwerten im Vergleich zu einer Auswertung mit Target-Übereinstimmung (*Tabelle 14*). *Hypothese 1c* wird angenommen. Das Ziel der Profilauswertung, ein Scoring zu generieren, bei dem individuelle Antworttendenzen eine untergeordnete Rolle spielen, wurde somit erreicht. Dies wurde auch in der Korrelation zwischen ViTEIP-V2 und TEMINT (*Tabelle 15*) sichtbar, die nunmehr nicht mehr im negativ signifikanten Bereich liegt, sondern sich als positive Korrelation ($r \leq 0.16^*$) darstellt. Der geringe Zusammenhang zwischen den beiden Emotionale-Intelligenz-Tests entspricht zwar nach wie vor nur knapp den formulierten Erwartungen, ist aber dennoch inhaltlich besser nachvollziehbar als die zuvor gefundenen Korrelationsmuster. Wie in Kapitel 1.4 vorgestellt, berichteten auch andere Autoren von nur geringen Korrelationen zwischen fähigkeitsbasierten Emotionale-Intelligenz-Tests,

wie beispielsweise Ciarrochi et al. (2003) oder Choi und Kluemper (2012). Es stellt sich die damit grundsätzliche Frage, ob alle Tests, die die Bezeichnung „Emotionale-Intelligenz-Test“ tragen, auch tatsächlich überlappende Bereiche desselben Konstrukts erfassen oder ob es sich hier um unterschiedliche Konstruktbereiche handelt. Erfasst der TEMINT, mit seiner knappen verbalen Beschreibung allgemeiner Alltagsszenarien dieselbe grundlegende Fähigkeit wie der ViTEIP mit den komplexen Situationen aus dem Pflegekontext? Die zum Teil deutlich unterschiedlichen Korrelationsmuster der beiden Emotionale-Intelligenz-Tests zu anderen Verfahren (*Tabelle 9*) könnten ein Hinweis sein, dass dies nicht der Fall ist.

Fazit: Profilauswertung. Insgesamt zeigten sich in der vorliegenden Arbeit für die Profilauswertung positive Ergebnisse. Mit diesem Scoring kamen individuelle Antworttendenzen in den Gesamttestergebnissen deutlich weniger zum Tragen als bei der Auswertung Target-Übereinstimmung. Die dort kritisierte Bevor- oder Benachteiligung von Probanden mit bestimmtem Antwortbias fand bei der Profilauswertung nicht statt. Kritisch zu sehen ist, dass durch die Zusammenfassung von Einzelitems zu Situationen ein Informations- und Varianzverlust entsteht. Obwohl die Probanden im ViTEIP-V2 insgesamt 99 Emotionen bearbeiteten, ergaben sich durch die Zusammenfassung zu Profilen nur 12 Einzelwerte, die in einen Gesamtwert einfließen, was die vergleichsweise geringen Reliabilitäten (maximal $\alpha \leq 0.65$) erklärt. Eine Verlängerung der Tests auf die ursprüngliche Itemanzahl von 99 bzw. 85 Items führte in den Stichproben bei Anwendung der Spearman-Brown-Korrekturformel überwiegend zu einer deutlichen Erhöhung der internen Konsistenz auf Werte $\alpha \geq 0.78$; eine nach wie vor geringe Reliabilität ergab sich für den ViTEIP-V2 in Stichprobe 2b. Da die „Items“ bei der Profilauswertung jedoch „Situationen“ sind, welche ein Konglomerat aus mehreren Items darstellen, ist diese Testverlängerung nur theoretisch, da die Zumutbarkeit eines solch langen Tests sehr fraglich ist. Eine Verlängerung des Tests zur Erhöhung der Reliabilität bietet sich somit praktisch nicht an; daher ist ein Einsatz der Profilauswertung vor allem aus testökonomischen Gründen nur bedingt zu empfehlen.

5.1.2.2 Übertragung der Signalentdeckungstheorie auf Emotionale-Intelligenz-Tests

Die Signalentdeckungstheorie (SET) als Scoringprozedur auf Emotionale Intelligenz zu übertragen, ist nach Kenntnisstand der Autorin eine Vorgehensweise, die außerhalb dieser Arbeit noch nicht zum Einsatz kam. Insofern sind die vorliegenden Ergebnisse als erste Hinweise zu verstehen, inwiefern dieser Auswertungsansatz für Emotionale-Intelligenz-Tests angewandt werden kann.

Die SET bietet sich als Scoring für Emotionale Intelligenz an, da die Unterscheidung in Rausch- und Signaldurchgänge auf den ViTEIP-V2 (bzw. TEMINT und andere Emotionale-Intelligenz-Tests) übertragen werden kann. In diesen Tests gibt es jeweils Durchgänge, in denen ein Reiz (d.h. eine

Emotion) bzw. ein Rauschen (d.h. keine Emotion) in Form eines Items dargeboten wird. Zudem zeigte sich in der vorliegenden Arbeit, dass im ViTEIP-V2 und TEMINT nicht nur die Fähigkeit zur Emotionsidentifikation bei der Bearbeitung der Items eine Rolle spielt, sondern auch individuelle Antwortneigungen berücksichtigt werden sollten. In der Theorie zur Signalentdeckung wird diesem Aspekt in Form des Kennwerts „Reaktionsneigung“³³ Rechnung getragen. Es wird in der SET angenommen, dass Probanden für ihre Antworten eine bestimmte Empfindungsintensität festlegen, ab der die Entscheidung von Rauschen zu Signal wechselt. Ein „idealer“ Proband ohne spezifische Reaktionsneigung ($\hat{c}=0$) weist eine Ausgeglichenheit zwischen Falschen Alarmen und Verpassern auf. Eine Veränderung der Reaktionsneigung wirkt sich auf den Anteil der Identifikationen, aber auch auf die Anteil der Falschen Alarme aus. Nachzuvollziehen ist diese Veränderung exemplarisch durch die Verschiebung der senkrechten, grauen Linie in *Abbildung 8* nach links oder rechts. An dieser Abbildung wird sichtbar, dass es in der SET aufgrund der Überschneidung der Rausch- und Signalverteilung quasi nicht möglich ist, das Antwortkriterium so zu setzen, dass ausschließlich Identifikationen und keine Falschen Alarme entstehen.

Deskriptive Kennwerte der Signalentdeckungstheorie. Anhand der Mittelwerte in *Tabelle 13* wurde deutlich, dass es den Probanden sowohl im ViTEIP-V2 als auch im TEMINT im Mittel gleich gut gelang, vorhandene Emotionen richtig zu identifizieren. Hier lagen die Werte für beide Tests innerhalb der einzelnen Stichproben in gleicher bzw. vergleichbarer Höhe (z.B. Stichprobe 1: beide Tests $M=0.81$ für Identifikation). Deutliche Unterschiede zeigten sich jedoch in der Leistung bezogen auf den Umgang mit den Null-Items (d.h. den Items mit Angaben der Originalperson „Emotion nicht vorhanden“). Hier fiel es den Probanden im Mittel im TEMINT erheblich leichter, die korrekte Antwort zu geben; der Mittelwert für Korrekte Zurückweisung lag in allen drei Stichproben $M \geq 0.95$. Im ViTEIP-V2 fielen diese Mittelwerte in den drei Stichproben deutlich geringer aus ($M \leq 0.60$). Werden zusätzlich zu den Mittelwerten die Korrelationen zwischen den Kennwerten der beiden Tests betrachtet, wird deutlich, dass Identifikation und Korrekte Zurückweisung in unterschiedlichem Maße die Korrelation zu Target-Übereinstimmung speisen (*Tabellen A-28 bis A-30, Anhang H*). Im ViTEIP-V2 (Stichprobe 1) korrelierte Identifikation zu $r=-0.34^{**}$ mit Target-Übereinstimmung und Korrekte Zurückweisung zu $r=0.82^{**}$; im TEMINT lagen die Korrelationen bei $r=0.85^{**}$ und $r=0.03$. Hier zeigte sich, dass mit der Auswertung Target-Übereinstimmung die hinter dem Gesamtwert liegenden Zusammenhänge verschleiert werden. Zwischen ViTEIP-V2 und TEMINT lagen die Korrelationen für den Kennwert Identifikation in den Stichproben zwischen $r \geq 0.43^{**}$ und $r \leq 0.49^{**}$ sowie für Korrekte Zurückweisung zwischen $r \geq 0.23^{**}$ und $r \leq 0.31^{**}$. Diese Unterscheidung bietet eine erste Erklärung, warum der Zusammenhang zwischen ViTEIP-V2 und TEMINT bei

³³ Für eine Beschreibung der verschiedenen Kennwerte sei auf Kapitel 4.4.2 verwiesen.

targetbasierter Auswertung unerwarteter Weise negativ ausfällt. Es erscheint daher sinnvoll, eine differenzierte Auswertung unter Berücksichtigung der beiden „Emotionsausprägungen“ (d.h. Emotion vorhanden vs. Emotion nicht vorhanden“) vorzunehmen.

Detektionsleistung. Die Detektionsleistung \hat{d}' gilt in der Signalentdeckungstheorie als ein valides Maß für die Erkennensleistung, da sowohl die Identifikations- als auch die Fehlalarmrate berücksichtigt wird. Sie stellt den Abstand zwischen den Mittelwerten der beiden Verteilungen dar: Je größer die Detektionsleistung eines Probanden ausfällt, desto weiter liegen die Signal- und Rauschverteilung auseinander und je besser gelingt die Unterscheidung zwischen Rauschen und Signal. Die Betrachtung der Detektionsleistung stellt damit eine mögliche targetbasierte Alternative zu den bisher in der Literatur vorgeschlagenen Scorings für Emotionale-Intelligenz-Tests dar. Die Untersuchung der Eignung erfolgte mit *Hypothese 1d*: „Die Detektionsleistung ist ein geeigneter Indikator, die Leistung des Probanden biasfrei abzubilden. Dies zeigt sich in einer signifikant geringeren Korrelation zwischen der Detektionsleistung und der „Häufigkeit der gewählten Antwortoptionen“ als bei Target-Übereinstimmung.“. Der mittlere Abstand zwischen den beiden Verteilungen war beim TEMINT ($\hat{d}' = 3.47$, Stichprobe 1) deutlich größer als beim ViTEIP-V2 ($\hat{d}' = 0.63$). Dieses Ergebnis zeigte, dass den Probanden im TEMINT die Differenzierung zwischen Rauschen (d.h. Angabe der Originalperson lautet „Emotion nicht vorhanden“) und Signal (d.h. Angabe der Originalperson lautet „Emotion schwach/mittel/stark vorhanden“) besser gelang als im ViTEIP-V2. Die Detektionsleistung stand in einem vergleichsweise geringen Zusammenhang mit dem Antwortbias, wie in *Tabelle 18* für beide Tests und in allen Stichproben zu sehen ist. Damit erfüllte dieser Kennwert die formulierte Erwartung; die *Hypothese 1d* wird angenommen.

Die Unterscheidung, die in der SET mit der Detektionsleistung als biasfreien Leistungsindikator und der Reaktionsneigung als Kennwert zum Antwortverhalten vorgenommen wird, zeigte sich auch in den vorliegenden Ergebnissen. Die Korrelationen der Kennwerte zum Antwortverhalten fielen zur Reaktionsneigung deutlich höher aus als zur Detektionsleistung (*Tabelle 18*). Inwiefern die Detektionsleistung tatsächlich einen Leistungsindikator darstellt, ist schwierig zu beurteilen. Es zeigten sich für den ViTEIP-V2 in Stichprobe 1 größtenteils die unter *Hypothese 1e* formulierten Korrelationsmuster zur konvergenten Validität (*Tabelle 20*): die Detektionsleistung korrelierte zu $r = 0.24^{**}$ mit Intelligenz und $r = 0.29^{**}$ mit dem EKF (Gesamt). In den anderen Stichproben und für den TEMINT bestätigten sich die Erwartungen jedoch nicht; hier lagen die Korrelationen teilweise im negativen Bereich. Zwischen ViTEIP-V2 und TEMINT korrelierten die Detektionsleistungen nicht signifikant mit $r = 0.05$ (Stichprobe 1 und 2a) und $r = 0.07$ (Stichprobe 2b).

Insgesamt ist die Detektionsleistung als Kennwert für Emotionale-Intelligenz-Tests eingeschränkt zu empfehlen. Vorteil des Scorings ist, dass der Kennwert weitestgehend biasfrei und somit unabhängig von Antworttendenzen der Probanden die Leistung in den Tests erfasst. Solange im Rahmen der Validitätsanalyse nicht hinreichend geklärt werden kann, was die Tests tatsächlich erfassen, bleibt der Einsatz dieses Scorings fraglich.

5.1.2.3 Die Kennwerte Emotionale Sensitivität und Emotionale Spezifität

Es fand eine Weiterentwicklung des Grundgedankens der Signalentdeckungstheorie in der Übertragung auf Emotionale Intelligenz statt und es wurde ein neues Scoring mit den Kennwerten Emotionale Sensitivität und Emotionale Spezifität entwickelt. Für die Items wurden zunächst die Übereinstimmungswerte mit dem Target berechnet. Alle Items, die von Originalpersonen mit „Emotion vorhanden“ kommentiert waren, wurden zu einem Gesamtwert „Emotionale Sensitivität“ aufsummiert sowie alle Items mit der Angabe der Originalperson „Emotion nicht vorhanden“ zum Gesamtwert „Emotionale Spezifität“. Vorteil dieser Kennwerte ist, dass durch die Verwendung von Übereinstimmungswerten vorhandene Varianzen beibehalten werden. Zudem zeigte sich anhand der vorliegenden Ergebnisse, dass eine getrennte Auswertung von sogenannten Null- und Nicht-Null-Items³⁴ sinnvoll ist.

Da die internen Konsistenzen von Emotionaler Sensitivität in allen drei Stichproben mit Werten zwischen $\alpha \geq 0.51$ und $\alpha \leq 0.63$ nur als moderat einzuordnen sind, wurde probeweise eine Alphasmaximierung an den Daten der Stichprobe 1 durchgeführt und an Stichprobe 2a und 2b validiert. Durch Entfernung von nicht-trennscharfen Items konnte eine Erhöhung der internen Konsistenz auf akzeptable Werte $\alpha \geq 0.70$ erreicht werden. Für zukünftige Anwendungen des ViTEIP könnte die Anwendung alphasmaximierter Skalen sinnvoll sein. Bevor eine entsprechende Itemselektion vorgenommen wird, sollte der ViTEIP-V2 jedoch in der aktuellen Version an weiteren Stichproben zum Einsatz kommen.

Überprüfung der Kennwerte in Bezug auf den Antwortbias. Die Untersuchung zu den Auswirkungen des Antwortbias erfolgte bei diesem Auswertungsansatz in anderer Weise als bei den vorangegangenen Kennwerten. Durch die Aufteilung in Emotionale Sensitivität und Emotionale Spezifität war es inhaltlich sinnvoll, nicht auf die Höhe der Korrelation, sondern auf die Richtung (d.h. das Vorzeichen) der Korrelation zu schauen. Konkret bedeutete dies, dass sich eine positive Korrelation zwischen der Häufigkeit der gewählten Antworten „Emotion vorhanden“ und Emotionaler Sensitivität sowie zwischen „Emotion nicht vorhanden“ und Emotionaler Spezifität

³⁴ Null-Items= Items, bei denen die Antwort der Originalperson „Emotion nicht vorhanden“ lautet. Nicht-Null-Items= Items, bei denen die Antwort der Originalperson „Emotion schwach/mittel/stark vorhanden“ lautet.

zeigen sollte. Vor allem für letztgenannten Zusammenhang traf dies sehr deutlich mit Korrelationen über $r \geq 0.86^{**}$ zu. Für den TEMINT fielen die Korrelationen bei den Nicht-Null-Items ähnlich hoch aus; beim ViTEIP-V2 lagen die Korrelationen etwas niedriger. Die *Hypothese 1d* wird in diesem Kontext angenommen. Insgesamt ist die Bezeichnung „biasfrei“ für die beiden neuen Kennwerte nicht ganz korrekt, da der Bias nicht herausgerechnet wird, sondern durch die Trennung in Null-Items und Nicht-Null-Items eine differenzierte Auswertung vorgenommen wird. Probanden mit einem spezifischen Ankreuzverhalten haben bei dieser Auswertungsform die Chance, zumindest in einem Teilbereich eine gute Leistung zu erzielen. Aus den getrennten Kennwerten kann eine Rückmeldung generiert werden, ob ein Proband eher Schwierigkeiten darin hat, vorhandene Emotionen (Emotionale Sensitivität) zu identifizieren oder ob die Einschätzung abwesender Emotionen (Emotionale Spezifität) problematisch ist. Die Frage danach, was ein Testergebnis aussagt, lässt sich im Gegensatz zur „Detektionsleistung“, die aufgrund der komplexeren Berechnung inhaltlich schwieriger zu interpretieren ist, für Emotionale Sensitivität und Emotionale Spezifität auch für Laien verständlich beantworten: je höher der Wert, umso höher die Übereinstimmung mit den Antworten der Originalpersonen, unterschieden in vorhandene und nicht vorhandene Emotionen.

Konvergente Validität. Die Korrelationen zur Überprüfung der konvergenten Validität fielen für die beiden Kennwerte sehr unterschiedlich aus. Für Emotionale Spezifität ergaben sich vor allem in Stichprobe 1 für den ViTEIP-V2 einige signifikante Korrelationen in erwarteter Richtung. Für die anderen Stichproben, den TEMINT sowie Emotionale Sensitivität beim ViTEIP-V2 waren die Zusammenhänge überwiegend nicht signifikant. Insgesamt führen die Ergebnisse daher zu einer Ablehnung der *Hypothese 1e* für die neuen Kennwerte.

Warum fielen die Korrelationen des ViTEIP-V2 zu anderen Verfahren in erster Linie bei Emotionaler Spezifität und nicht bei Emotionaler Sensitivität signifikant aus? Am deutlichsten war der Unterschied in der Korrelation zu verbaler Intelligenz zu sehen. Zu Emotionaler Spezifität lag diese Korrelation erwartungsgemäß bei $r=0.31^{**}$, wohingegen sie für Emotionaler Sensitivität nicht signifikant war. Ein weiteres Beispiel ist die EKF-Facette „Erkennen Emotionen anderer“ mit einer Korrelation von $r=0.20^*$ zu Emotionaler Spezifität und einer Korrelation von $r=-0.19^*$ zu Emotionaler Sensitivität. Eine mögliche Erklärung könnte sein, dass die Anforderung des ViTEIP-V2, aufgrund der Situationsbeschreibung eine Einschätzung über Emotionen Anderer vorzunehmen, Probanden große Schwierigkeiten bereitete und die Umsetzung der Instruktion „wie fühlt sich Person xy am Ende der Situation“ nicht gelingt. Vielleicht können Probanden gerade noch das „Nicht-Vorhandensein“ von Emotionen rational lösen, aber die Intensität von Emotionen bei anderen einzuschätzen ist zu schwierig. Gegebenenfalls gelingt dies den Probanden nur über den

gedanklichen Umweg der Übertragung „Was würde ich fühlen?“ und dies führt zu heterogenen Ergebnissen. In den Freitexten zur Kommentierung des ViTEIP und TEMINT meldeten die Probanden zurück, dass sie es als schwierig empfanden, mit den präsentierten Informationen auf Emotionen zu schlussfolgern. Sie formulierten den Wunsch nach mehr Informationen oder einer Präsentation der konkreten Situationen im Video. Dieser Aspekt könnte in zukünftigen Untersuchungen eruiert werden, indem die Informationsmenge in den Items/Situationen systematisch variiert und die Auswirkungen auf die Antworten der Probanden untersucht wird. Wie viel Information wird benötigt, um eine Einschätzung der Emotionen anderer vorzunehmen? Ab wann führen zusätzliche Informationen nicht mehr zu einer besseren Leistung? Mit diesen Erkenntnissen könnte möglicherweise auch die Güte des Kennwerts Emotionale Sensitivität erhöht werden.

Roberts et al. (Roberts, Zeidner, & Matthews, 2001) kritisierten die geringen Korrelationen einzelner MEIS-Subskalen zu Intelligenz, woraufhin Mayer et al. auf die unterschiedliche kognitive Sättigung der Aufgaben verweisen: *“... our model supposes that the third branch, understanding, is most cognitive and should have the highest relation to abstract reasoning. That, in fact, is what the data show; the third branch does correlate most highly with IQ ...”* (Mayer J. D., Salovey, Caruso, & Sitarenios, 2001, p. 235). Diese Argumentation lässt sich theoretisch auf die vorliegenden Daten übertragen, bestätigte sich empirisch aber nicht. Dass Emotionale Sensitivität und Emotionale Spezifität gegebenenfalls unterschiedliche Anforderungen an die Probanden stellen, ist nachvollziehbar. Anzunehmen ist allerdings, dass sich die Aufgabe, vorhandene Emotionen zu identifizieren komplexer gestaltet als der Umgang mit nicht vorhandenen Emotionen. Damit würde erwartet, dass sich für Emotionale Sensitivität der engere Zusammenhang mit Intelligenz finden sollte, was in der vorliegenden Arbeit nicht der Fall ist.

In Rosts (2009) Werk über Intelligenz findet sich eine Vorstellung des Intelligenzmodells von Brandstätter, in dem interessanterweise eine ähnliche Unterteilung vorgenommen wird, wie sie in der vorliegenden Arbeit vorgeschlagen wird. Intelligenz zeigt sich nach Brandstätter darin, ob Personen vorhandene Muster korrekterweise als solche identifizieren oder ob Personen „Unordnung“ auch als diese erkennen und nicht fälschlicherweise das Vorhandensein einer Systematik hineininterpretieren. Das Modell von Brandstätter ist jedoch wenig verbreitet und konzeptionell weiter entfernt von etablierten Intelligenzmodellen. Daher soll die Existenz dieses Ansatzes nicht überstrapaziert oder dazu herangezogen werden, um als Beleg zu dienen, dass Emotionale Intelligenz tatsächlich eine Form von Intelligenz darstellt; dieser Beleg sollte in erster Linie über Studien dergestalt erfolgen, wie sie beispielsweise von MacCann et al. (MacCann, Joseph, Newman, & Roberts, 2014) durchgeführt wurden. Dennoch ist es gut nachvollziehbar, dass das Erkennen von Mustern und Systematik in der Bewältigung von Alltagsproblemen eine förderliche

Fähigkeit darstellt, und ist als Grundgedanke in Kombination mit dem Signalentdeckungsansatz auch auf den Umgang mit Emotionen übertragbar. Die hier vorliegenden Ergebnisse zu diesem neuen Scoringansatz stellen eine gute Basis für weitere Überlegungen und Untersuchungen dar.

Über die beiden Emotionale-Intelligenz-Tests VITEIP-V2 und TEMINT hinweg war die Korrelation der jeweils gleichen Kennwerte erwartungsgemäß positiv; die beiden Kennwerte zueinander standen hingegen in einem negativen Zusammenhang. Grundsätzlich entspricht diese negative Korrelation weder der Annahme eines eindimensionalen Emotionale-Intelligenz-Konstrukts (Mayer, Caruso, & Salovey, 2000) noch der Annahme der positiven Mannigfaltigkeit von Intelligenz (Guttman & Levy, 1991). Dass es innerhalb eines Emotionale-Intelligenz-Tests zu negativen Korrelationen kommt, war jedoch nicht nur in der vorliegenden Untersuchung der Fall. Ein vergleichbares Ergebnis zeigte sich bei Follesdal und Hagtvet (2009), wo sich im MSCEIT zwei distinkte Faktoren für vorhandene und abwesende Emotionen in der Aufgabe „Gesichter“ ergaben. Die wenigen im Test integrierten Items zu „vorhandene Emotionen“ korrelierten nur schwach oder sogar negativ zu den Items von „abwesende Emotionen“. Auch Bänziger et al. (Bänziger, Grandjean, & Scherer, 2009) fanden in ihren Untersuchungen mit der MERT einen visuellen und einen vokalen Faktor der Emotionserkennung, die weitestgehend unabhängig voneinander waren. Wird Emotionale Intelligenz in den Zusammenhang des Signalentdeckungsansatzes gestellt, kann eine negative Korrelation jedoch nachvollziehbar interpretiert werden, da es nur schwer möglich ist, Sensitivität und Spezifität gleichermaßen zu erhöhen. Exemplarisch bedeutet dies, dass ein Proband sein Ergebnis in Emotionaler Sensitivität dadurch erhöhen könnte, indem er insgesamt über den Test hinweg häufiger „Emotion vorhanden“ antwortet. Damit steigt jedoch auch die Wahrscheinlichkeit, dies bei Items zu tun, bei denen die Antwort „Emotion nicht vorhanden“ lautet und somit in den Gesamtwert von Emotionaler Spezifität einfließt. Nur einem „idealen“ Probanden gelänge es durch gezieltes (korrektes!) Antworten, seine Ergebnisse in dem einen Kennwert zu verbessern, ohne dass dies zulasten des anderen Kennwerts ginge.

Fazit: Emotionale Sensitivität und Emotionale Spezifität. Wie beschrieben ist es inhaltlich und aufgrund der Datenlage plausibel, eine Unterscheidung vorzunehmen, die sich auf den Umgang mit vorhandenen bzw. mit nicht vorhandenen Emotionen bezieht. Emotionale Sensitivität und Emotionale Spezifität bilden damit vermutlich zwei unterschiedliche Bereiche eines Emotionale-Intelligenz-Konstrukts ab. Mit einer Aufteilung der Items zu zwei Kennwerten konnten die Korrelationsmuster besser erklärt werden als mit nur einem Kennwert: die neu generierten Kennwerte wiesen zum Teil zu anderen Testwerten unterschiedliche Richtungen der Vorzeichen auf und haben damit unterschiedlichen Validitätsansprüche (*Tabelle 20 und A-26, A-27 in Anhang H*).

Dass verschiedene Aspekte von Emotionaler Intelligenz in unterschiedlichen Zusammenhängen zu anderen Verfahren stehen, fand sich auch in weiteren Untersuchungen. Bei Libbrecht und Lievens (2012) korrelierte die Facette „Emotionale Regulation“ mit interpersonellen akademischen Leistungen, die Facette „Emotionales Verständnis“ hingegen nicht. Auch Fiori und Antonakis (2011) regten an, für den MSCEIT aufgrund unterschiedlicher Korrelationen zu Persönlichkeitsmerkmalen die jeweiligen Skalenwerte zu verwenden; ein Einsatz des Gesamtwertes verschleierte die wahren Zusammenhänge. Dieses Ergebnis wird durch die vorliegende Arbeit bestätigt, da von der Anwendung der Gesamtwerte, die sich bei den ‚klassischen‘ Scorings ergeben, abzuraten ist und die Verwendung der differenzierenden Kennwerte Emotionale Sensitivität und Emotionale Spezifität empfohlen wird.

Um Alltags- und berufliche Situationen erfolgreich zu bewältigen, sind beide Fähigkeiten gleichermaßen erforderlich. Zum einen ist es von Bedeutung, auf vorhandene oder entstehende Emotionen innerhalb einer Situation angemessen reagieren zu können. Um eine Reaktion einleiten zu können, ist es im ersten Schritt wesentlich, die Emotion seitens des Gegenübers korrekt zu identifizieren. Eine Reaktion auf vorhandene Trauer seitens des Interaktionspartners fällt in der Regel anders aus, als wenn die Annahme besteht, die vorherrschende Emotion sei Ärger. Zum anderen geht es um die Fähigkeit, aus Situationen (z.B. im Rahmen einer Begegnung, im Gespräch) richtig schlussfolgern zu können, dass eine oder mehrere Emotionen nicht vorhanden sind. Auch hier ist es entscheidend, mit dem Gegenüber adäquat umzugehen, und beispielsweise Emotionen anzunehmen, obwohl diese nicht vorliegen. Um zu verdeutlichen, wie sich eine Situation entwickeln kann, in der von falschen Schlussfolgerungen bezüglich der emotionalen Lage des Gegenübers ausgegangen wird, sei – mit einem kleinen Augenzwinkern – auf die allseits bekannte „Geschichte mit dem Hammer“ von Watzlawick (2004) verwiesen (siehe *Anhang K*).

Um zu prüfen, ob die Unterscheidung Emotionale Sensitivität und Emotionale Spezifität praktische Relevanz besitzt, könnte eine Untersuchung an spezifischen Probandengruppen erfolgen, beispielsweise an Personen mit diagnostizierter Alexithymie oder sozialer Phobie. Beide Personengruppen würden vermutlich eher geringe Gesamtwerte in Emotionaler Intelligenz erreichen. Die Unterscheidung in Emotionale Sensitivität und Emotionale Spezifität würde ein differenzierteres Ergebnismuster ermöglichen. Zu erwarten wäre, dass Personen mit Alexithymie aufgrund der eher geringen Kompetenz Emotionen bei anderen und bei sich selbst zu identifizieren, hohe Werte in Emotionaler Spezifität und geringe Werte in Emotionale Sensitivität erreichten. Personen mit sozialer Phobie hingegen besitzen häufig aufgrund ihrer Sorge vor Bewertungen durch andere eine

hohe Aufmerksamkeit gegenüber kleinster körpersprachlicher oder emotionaler Signale³⁵. Hier würde die Ausprägung entgegengesetzt ausfallen. Insgesamt würden die Antwortkategorien zu „Emotion vorhanden“ häufiger gewählt als bei einer gesunden Stichprobe. In beiden Fallkonstellationen könnte die Unterscheidung von Emotionaler Sensitivität und Emotionaler Spezifität dazu genutzt werden, jeweils vorhandene Kompetenzen wertzuschätzen und gezielt an Defiziten zu arbeiten. Die Einsatzmöglichkeit bzw. Differenzierungskraft bei klinischen Gruppen stellt natürlich ein Extrembeispiel dar und der Nutzen und die Anwendbarkeit des ViTEIP-V2 müssten sich in diesem Zusammenhang erst noch zeigen. Dennoch kann es aus Sicht der Autorin sowohl in klinischen als auch nicht-klinischen Kontexten hilfreich sein, die Unterscheidung in Emotionale Sensitivität und Emotionale Spezifität als Basis für Beratung und Coaching zu nutzen.

5.1.3 Fazit: Auswertungsstrategien

Die vorliegende Arbeit zielte darauf ab, unterschiedliche Scorings systematisch zu untersuchen und aus den Ergebnissen abzuleiten, welche Scorings sich für die Anwendung bei Emotionale-Intelligenz-Tests am besten eignen. Aus den Untersuchungen ergibt sich aus Sicht der Autorin die Empfehlung, auf targetbasierte Auswertungen zurückzugreifen: Neben grundlegenden kritischen Einwänden bezüglich des Konsensscorings sprechen eine höhere Varianz und Reliabilität sowie etwas deutlichere Validitätsbelege für den Einsatz des Targetscorings.

Dem großen Kritikpunkt der negativen Korrelation zwischen ViTEIP-V2 und TEMINT bei targetbasierten Auswertungen konnte mit den neuen Scorings entgegengewirkt werden. Alle neuen Auswertungsansätze reduzierten den Einfluss des vorliegenden Antwortbias deutlich, was ein eindeutiger Vorteil gegenüber den bisher bekannten targetbasierten Auswertungen darstellt. Aus den Korrelationsmustern zu anderen Verfahren konnte jedoch nicht uneingeschränkt gefolgert werden, dass mittels ViTEIP-V2 und TEMINT tatsächlich Komponenten Emotionaler Intelligenz erfasst werden. Die als insgesamt kritisch zu beurteilenden Validitäten der beiden eingesetzten Emotionale-Intelligenz-Tests (unabhängig von den Scorings) werden in Kapitel 5.4 diskutiert.

Die Profilauswertung ergab insgesamt positive Ergebnisse und eignet sich grundsätzlich als Scoringprozedur für Emotionale-Intelligenz-Tests. Die eher moderaten Reliabilitäten in den vorliegenden Stichproben sind vermutlich auf die geringe Anzahl an Items zurückzuführen, die in die Kennwerte einfließen. Da aufgrund des Auswertungsprinzips über Situationen von einer Testverlängerung zur Steigerung der Reliabilität abzusehen ist, ist ein Einsatz dieser Scoringprozedur unter Berücksichtigung der Testökonomie eher fraglich.

³⁵ Bei sozialer Angst bzw. Phobie ist zu berücksichtigen, dass es neben der oben beschriebenen Personengruppe auch Personen gibt, die aufgrund einer geringen Ausprägung sozialer Kompetenzen „berechtigte“ soziale Ängste besitzen.

Mit der Anwendung der Signalentdeckungstheorie bei Emotionale-Intelligenz-Tests wurde ein grundlegend neuer Auswertungsansatz vorgestellt. Der Kennwert „Detektionsleistung“ bildete die Leistung eines Probanden im ViTEIP-V2 oder TEMINT so ab, dass der individuelle Antwortbias nur eine geringe Rolle spielt. Wie schon im Kontext anderer Auswertungsansätze überwogen nicht signifikante Korrelationen und die Frage nach der Validität der Tests blieb unbeantwortet. Bezogen auf den aktuellen Wissensstand ist die Verwendung des Kennwertes nur für Forschungszwecke zu empfehlen, gleichwohl ein Einsatz in diesem Kontext stark zu befürworten ist, um zusätzliche Erkenntnisse zu gewinnen.

Die beiden Kennwerte Emotionale Sensitivität und Emotionale Spezifität wurden aus der Signalentdeckungstheorie und Target-Übereinstimmung abgeleitet, beziehen sich auf ein targetbasiertes Grundprinzip und reduzieren gleichzeitig die Nachteile, die sich aus den anderen targetbasierten Ansätzen ergeben. Mit der getrennten Auswertung von Null- und Nicht-Null-Items wurden die Bevorzugungen bzw. Benachteiligungen, die sich aus einem spezifischen Ankreuzverhalten eines Probanden ergaben, deutlich minimiert. Zudem zeigte sich in den Korrelationen zu anderen Verfahren, dass Emotionale Sensitivität und Emotionale Spezifität unterschiedliche Validitätsansprüche aufweisen; diese wurden bei allen anderen Scorings verschleiert. Die Autorin empfiehlt daher, die getrennte Auswertung für die beiden vorliegenden Emotionale-Intelligenz-Tests zu nutzen und weitere Forschungsbemühungen anzustrengen, diesen neuen Scoringansatz zu optimieren und zu validieren.

5.2 Antwortformate

Untersucht wurde, inwiefern sich die Ergebnisse im ViTEIP-V2 auf Stichprobenebene verändern, wenn im Testheft ein anderes Antwortformat dargeboten wird. Die vorgenommenen Änderungen bezogen sich darauf, wie die einzuschätzenden Emotionen präsentiert wurden und in welcher Form die Intensitätseinschätzung durch die Probanden zu erfolgen hatte. Bei dem Antwortformat „Ankreuzen“ wurde eine Liste von Emotionen schriftlich dargeboten und die Probanden wurden gebeten, diese auf einer Likertskala bezüglich der Intensität einzuschätzen. Bei „Zuordnen“ wurden zunächst alle Emotionen in einem Bearbeitungsfeld schriftlich aufgeführt. Aufgabe der Probanden war es, die Emotionen aus diesem Feld heraus weiteren Feldern entsprechend der vorhandenen Intensität zuzuordnen. Alle Emotionen, die im ursprünglichen Bearbeitungsfeld verbleiben, wurden der Intensität „Emotion nicht vorhanden“ zugeordnet.

Idealerweise würde eine Untersuchung der Auswirkungen eines anderen Antwortformates über einen intraindividuellen Vergleich erfolgen, indem jeder Proband jeweils beide Versionen bearbeitet

und die Ergebnisse auf Personenebene verglichen werden. Aus organisatorischen und zeitökonomischen Gründen war dies jedoch nicht möglich. Aus dieser Einschränkung ergibt sich, dass die Antwortformate über Vergleiche der Ergebnisse der Teilstichproben 2a und 2b (Altenhilfeschüler) untersucht wurden. Den Probanden der Teilstichprobe 2a wurde der ViTEIP-V2 mit dem Antwortformat „Ankreuzen“ dargeboten, der Teilstichprobe 2b mit „Zuordnen“. Beide Teilstichproben können als vergleichbar bezüglich demographischer Variablen (siehe *Tabelle 5*), der Ergebnisse in anderen eingesetzten Verfahren (*Tabelle A-5, Anhang D*) sowie der Akzeptanz des Verfahrens (*Tabelle A-31, Anhang I*) angesehen werden. Unterschiede, die sich auf Stichprobenebene im ViTEIP-V2 ergeben, lassen sich mit großer Wahrscheinlichkeit auf das veränderte Antwortformat zurückführen.

Zentrale Kennwerte. In der vorliegenden Untersuchung zeigte sich, dass eine Veränderung des ViTEIP-Antwortformates unter Anwendung der ‚klassischen‘ Scorings Target-Übereinstimmung, Target-Treffer, Konsens-Anteil und Konsens-Modal zu deutlichen Änderungen der Ergebnisse auf Ebene der Teilstichproben führte. Die Unterschiede, die sich bei der Anwendung der Profilauswertung und der Signalentdeckungstheorie ergeben, lagen hingegen im geringen Bereich. Die *Hypothese 2a* „Eine Veränderung von „Ankreuzen“ zu „Zuordnen“ bringt keine wesentlichen Änderungen in den zentralen Kennwerten mit sich“ wird für die vier Scorings Target-Übereinstimmung, Target-Treffer, Konsens-Anteil und Konsens-Modal zurückgewiesen; für die neuen Scorings mittels Profilauswertung und Signalentdeckungstheorie angenommen.

Bei den ‚klassischen‘ Scorings erreichte die Teilstichprobe 2b mit dem Antwortformat „Zuordnen“ höhere Mittelwerte als die Teilstichprobe 2a, in der der ViTEIP-V2 mit Antwortformat „Ankreuzen“ bearbeitet wurde (*Tabelle 21*); die Unterschiede in den Mittelwerten äußerten sich in mittleren bis großen Effektstärken zwischen $d=0.34$ und $d=0.81$. Da sich die Stichproben 2a und 2b in Bezug auf andere Stichprobenkennwerte oder Ergebnisse in weiteren Verfahren nicht unterscheiden, ist davon auszugehen, dass tatsächlich das geänderte Antwortformat und nicht eine höhere Ausprägung in Emotionaler Intelligenz innerhalb der Stichprobe ursächlich ist. Die Veränderung des Antwortverhaltens auf Stichprobenebene ist in dem Balkendiagramm der *Abbildung 11* deutlich zu erkennen. Vor allem der hohe hellblaue Balken bei „Emotion nicht vorhanden“ verdeutlicht, dass diese Kategorie in Stichprobe 2b im Mittel häufiger gewählt wurde als von Stichprobe 2a bei „Ankreuzen“. Beim Antwortformat „Zuordnen“ neigten Probanden eher dazu, Items der Kategorie „Emotion nicht vorhanden“ zuzuordnen“, was ihnen aufgrund der vergleichsweise hohen Anzahl an Items mit Angaben der Originalperson „Emotion nicht vorhanden“ zu Gute kam, da – zumindest bei targetbasierten Auswertungen – eine höhere Anzahl an „Emotion nicht vorhanden“ überwiegend zu besseren Testergebnissen führt.

Das vorgegebene Antwortformat wirkte sich nicht nur auf den Mittelwert der Stichprobe aus, sondern es zeigten sich in Kombination von Antwortformat und Auswertungsstrategie verschiedene Tendenzen. Wurden targetbasierte Scorings verwendet, fielen die Varianz und die Reliabilität bei „Ankreuzen“ größer als bei „Zuordnen“ aus. Bei konsensbasierten Scorings waren Varianz und Reliabilität bei „Zuordnen“ größer. Auch wenn im vorangegangenen Kapitel die Auswirkungen eines anderen Scorings als eher gering eingeschätzt wurden, ergab sich, dass die Entscheidung für oder gegen eine Auswertungsstrategie Konsequenzen für andere methodische Bereiche mit sich bringen kann. Dass sich, wie im vorliegenden Fall, Testergebnisse maßgeblich verändern, wenn ein anderes Antwortformat dargeboten wird bzw. ein anderes Antwortformat und ein bestimmtes Scoring zum Einsatz kommen, verweist auf die testtheoretisch ungünstige Situation der Existenz eines großen Varianzanteils, der nicht auf das Konstrukt, sondern auf die Methode zurückzuführen ist (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003).

Unter Anwendung der ‚neuen‘ Scorings bestätigte sich die *Hypothese 2a*; diese Ansätze reduzieren den Einfluss, den das unterschiedliche Antwortformat auf die Testleistung ausmacht. Wurde die Profilauswertung angewandt, fielen die Mittelwertunterschiede zwischen den Stichproben deutlich geringer aus: die Effektstärke bei einer Auswertung über Profile lag nur bei $d=0.02$. Bei der Anwendung der Signalentdeckungstheorie für die Detektionsleistung lag die Effektstärke bei $d=0.01$ (Tabelle 21). Die Kennwerte Emotionale Spezifität und Emotionale Sensitivität der getrennten Auswertung nach Null-Items (d.h. Angabe der Originalperson: Emotion nicht vorhanden) und Nicht-Null-Items (d.h. Angabe der Originalperson: Emotion vorhanden) werden zusammen betrachtet, da diese erst in der Kombination eine sinnvolle Information über eine Testleistung vermitteln. Bei dieser Auswertungsform wurde deutlich, dass die beiden Antwortformate bei den Probanden unterschiedliches Ankreuzverhalten begünstigten, was sich darin zeigte, dass bei „Ankreuzen“ (Teilstichprobe 2a) der höhere Mittelwert bei Emotionaler Sensitivität und bei „Zuordnen“ (Teilstichprobe 2b) der höhere Mittelwert bei Emotionaler Spezifität lag.

Insgesamt zeigte sich auf Stichprobenebene, dass im ViTEIP-V2 je nach Antwortformat verschiedene Antwortkategorien unterschiedlich häufig gewählt wurden. Das Antwortformat „Zuordnen“ scheint einen Aufforderungscharakter zu besitzen, mehr Emotionen mit der Intensität „nicht vorhanden“ zu beantworten. Auch wenn dieses Ergebnis im Zusammenhang mit dem ViTEIP-V2 für die Stichprobe 2b positiv ist und zu höheren Testergebnissen führte, fällt die Gesamtbewertung kritisch aus, da hier von sogenannten reaktiven Effekten auszugehen ist. Als reaktive Effekte werden allgemein methodenbedingte Effekte bezeichnet, welche durch das spezielle Verfahren der Datenerhebung bedingt sind und nicht durch das Merkmal selbst. Auch in anderen Untersuchungen führten veränderte Antwortformate zum Auftreten von Methodeneffekten. Batista-Foguet und Kollegen

(Batista-Foguet, Saris, Boyatzis, Guillén, & Serlavós, 2009) untersuchten beispielsweise, inwiefern sich die Anzahl der Abstufungen einer Likertskala auf die Ergebnisse von Emotionale-Kompetenz-Tests auswirkt. Sie stellten fest, dass eine 11-stufige Skala zu höherer Reliabilität, Validität und geringerer Invalidität führte als eine 5-stufige Skala. Gleichzeitig berichteten sie von Methodeneffekten, die sie in ihren Daten identifizieren konnten und welche z.B. bei der Modellierung von Strukturgleichungsmodellen über einen separaten Methodenfaktor zu berücksichtigen sind. Bei MacCann und Roberts (2008) fanden sich bei einem Multiple-Choice-Format im Vergleich zu einem Ratingskalenformat eine engere Korrelation zu Intelligenz.

Ob es sich tatsächlich um einen methodischen Bias handelt oder die Ergebnisunterschiede von einer grundsätzlichen Instabilität des Konstrukts Emotionale Intelligenz herrühren, kann aufgrund des Studiendesigns (interindividueller Vergleich von Ergebnissen) aus den vorliegenden Daten nicht abschließend beantwortet werden. Intraindividuelle Vergleich würden hier Klarheit bringen. Werden jedoch die Ergebnisse, die sich unter Anwendung der ‚neuen‘ Scorings ergeben, für die Interpretation hinzugenommen, scheint sich zu bestätigen, dass es sich um einen messmethodischen Effekt handelt. Da bei der Profilauswertung und den aus der Signal-entdeckungstheorie abgeleiteten Kennwerten dieselbe Datengrundlage wie für die anderen Scorings verwendet wurde, müssten sich Effekte, die vom Konstrukt und nicht von der Methode stammen, auch auf diese Ergebnisse auswirken. Bei diesen Auswertungen fielen die Unterschiede zwischen den beiden Teilstichproben deutlich geringer aus. Untersuchungen im MTMM-Design könnten mittels getrennter Ermittlung der trait- und methodenbedingten Varianz dabei helfen, weitere Informationen über Auswirkungen der Datenerhebungsmethode auf die Ergebnisse zu erhalten.

Korrelation zu anderen Verfahren. Die *Hypothese 2b* wurde bestätigt: es zeigten sich keine signifikanten Unterschiede in den Korrelationen zu anderen Verfahren. Es ergab sich ein ähnliches Bild wie bei der Untersuchung der verschiedenen Scorings: eine Veränderung des Antwortformats führte zu einigen Änderungen in den zentralen Kennwerten und zu keinen grundlegenden Änderungen in den Korrelationsmustern. Da in der vorliegenden Untersuchung in den Korrelationsmatrizen vorwiegend nicht signifikante Korrelationen zu finden waren, sollte dieses Ergebnis jedoch nicht überbewertet werden. Wie in Kapitel 5.4 diskutiert wird, stehen aufgrund der vorherrschenden Nullkorrelationen überzeugende Belege für die Validität des ViTEIP-V2 aus.

5.2.1 Fazit: Antwortformate

Aus der vorliegenden Arbeit ergibt sich die Empfehlung der Autorin, für den ViTEIP-V2 auf das Antwortformat „Ankreuzen“ zurückzugreifen. Der hohe Aufforderungscharakter bezüglich der

Nutzung der Antwortkategorie „Emotion nicht vorhanden“, den das Antwortformat „Zuordnen“ im Vergleich zu „Ankreuzen“ auf die Probanden zu haben scheint, verzerrt die Abbildung der tatsächlichen Leistungsfähigkeit. Auch die mittleren Wahlen der Antwortkategorien in Stichprobe 1, die in *Abbildung 11* nicht dargestellt wurden, jedoch vergleichbar zu Stichprobe 2a sind, bestätigen diese Vermutung. Bei Anwendung der ‚neuen‘ Scorings verringert sich zwar deutlich der Einfluss, den das Antwortformat auf die Testergebnisse hat, doch unter Berücksichtigung der oben genannten Ergebnisse sei auch hier die Verwendung von „Ankreuzen“ zu favorisieren.

5.3 Vergleich der Präsentationsmodi

Für den ViTEIP wurde ein Präsentationsformat gewählt, bei dem die Probanden ein kurzes Video der Originalperson bei einer neutralen Tätigkeit in ihrer Umgebung zu sehen bekommen. Parallel dazu werden von einer Sprecherin verschiedene Angaben zur Person und Situation vorgetragen. Ziel dieser Darstellung ist es, eine möglichst realitätsnahe Repräsentation der relevanten Ereignisse zu erhalten: Auch im Alltag besitzen Pflegekräfte in der Regel einige Hintergrundinformationen zu Person und Umgebung; das tatsächliche emotionsauslösende Ereignis wird jedoch meistens nicht direkt miterlebt, sondern Pflegenden erfahren erst im Nachhinein von der Begebenheit und müssen daraufhin auf die aktuelle Befindlichkeit der Patienten eingehen. Die *Hypothesen 3a bis 3d* wurden anhand des Vergleichs einer video- und einer textbasierten Forschungsversion des ViTEIP beurteilt. Die Ergebnisse zu den beiden Präsentationsmodi sind als erste Tendenz zu verstehen, da beide Testversionen nur aus fünf Situationen bestehen. Es wurden die Ergebnisse unter Anwendung der Scorings Target-Übereinstimmung, Target-Treffer, Konsens-Anteil und Konsens-Modal betrachtet. Aufgrund der geringen Itemanzahl in der Forschungsversion wurde von einer Auswertung über Profile sowie die Übertragung des Signalentdeckungsansatzes, bei dem zwischen Null- und Nicht-Null-Items unterschieden wird, abgesehen.

Korrelation zu Intelligenz. Die *Hypothese 3a* „Die videobasierte Version des ViTEIP korreliert in geringerem Ausmaß mit verbaler Intelligenz als die textbasierte Version“ wird abgelehnt, da die Korrelation zu verbaler Intelligenz bei beiden Versionen bei identischen Scorings jeweils in vergleichbarer Höhe im Bereich von $r \geq 0.23^{**}$ bis $r \leq 0.36^{**}$ für targetbasierte Auswertungen bzw. $r \geq 0.13$ bis $r \leq 0.18$ für konsensbasierte Auswertungen lag (*Tabelle 22*). Die Unterschiede in den Korrelationen zwischen den Versionen waren nicht signifikant. Die von Lievens und Sackett (2006) identifizierte geringere Korrelation mit kognitiven Fähigkeiten für die videobasierte Version eines Situational Judgment Tests im Vergleich zu einer schriftlich dargebotene Version, konnte nicht repliziert werden.

Für die Interpretation ist es hilfreich sich zu vergegenwärtigen, wie die text- und videobasierten Versionen des ViTEIP aufgebaut sind. Wie oben beschrieben, sehen Probanden bei der Videoversion ein Video einer Originalperson bei einer neutralen Tätigkeit. Parallel dazu werden verschiedene Hintergrundinformationen mündlich von einer Sprecherin vorgetragen; diese Informationen werden bei der Textversion ausschließlich schriftlich (ohne Video, ohne Sprecherin) dargeboten. Ob diese Informationen mündlich mit Video oder schriftlich präsentiert wurden, scheint für den Zusammenhang zu verbaler Intelligenz im ViTEIP irrelevant zu sein. Als Erklärung kann die überwiegend studentische Stichprobenzusammensetzung dienen. Hiermit handelt es sich um eine vergleichsweise leistungsstarke Stichprobe, die es aufgrund der Anforderungen im Studium gewohnt ist, sich mit verbalen Inhalten sowohl in gesprochener als auch schriftlicher Form gleichermaßen auseinanderzusetzen und somit grundsätzlich mit beiden Präsentationsmodi zurechtkommt.

Für die Höhe der Korrelation zwischen ViTEIP und verbaler Intelligenz war vielmehr von Bedeutung, welches Scoring zum Einsatz kam. Targetbasierte Auswertungen führten zu einer signifikant-positiven Korrelation, bei konsensbasierten Auswertungen war der Zusammenhang positiv, aber nicht signifikant. Zudem wurde deutlich, dass die Anwendung unterschiedlicher Scorings nicht nur Auswirkung auf die Korrelation zu verbaler Intelligenz, sondern auch auf das Gesamtergebnis im ViTEIP hatte: Targetbasierte Auswertungen ergaben höhere Mittelwerte in der Textversion; konsensbasierte Auswertungen in der Videoversion. Die Analyse der Freitexte der Probanden zu den beiden Versionen führt zu der Annahme, dass die Videoversion von den Probanden als nicht hilfreich für die Itembearbeitung empfunden wurde. Die Probanden kritisierten, dass die neutrale Darstellung der Personen ohne situationsbezogene Mimik und Gestik mehr irritierten denn halfen. Einige Probanden gaben an, die Videos hätten sie sogar vom Inhalt abgelenkt. Die Videos führten die Probanden somit möglicherweise in eine Richtung, die nicht der Targetantwort entspricht, wodurch die Probanden mittels targetbasierter Auswertung ein schlechteres Testergebnis erhalten. Für konsensbasierte Auswertungen ist ausschließlich die Übereinstimmung innerhalb der Referenzgruppe von Bedeutung; selbst wenn die Mehrheitsmeinung nicht auf die Targetantwort konvergiert, kann bei einer hohen Übereinstimmung innerhalb der Referenzgruppe ein hoher Testwert erreicht werden. Dies ist aus Sicht der Autorin eine grundlegende Schwäche des Konsensscorings.

Akzeptanz des Verfahrens. In Bezug auf die Akzeptanz des Verfahrens, für die zu erwarten war, dass die videobasierte Version im Vergleich zur textbasierten Version bei den Probanden zu höheren Akzeptanzwerten führt (Chan & Schmitt, 1997; Richman-Hirsch, Olson-Buchanan, & Drasgow, 2000), zeigten sich keine bedeutsamen Unterschiede. *Hypothese 3b* wird abgelehnt.

Damit zeigt sich auch der zweite erwartete Vorteil einer videobasierten Testung nicht: verbale Intelligenz ist für beide Formate im selben Ausmaß von Bedeutung und es gibt keine Unterschiede in Bezug auf die Akzeptanz. Wurden die schriftlichen Kommentare der Probanden bezüglich der Testversionen betrachtet, lässt sich dieses Ergebnis besser einordnen. Von den Probanden wurde bei der videobasierten Version des Tests einerseits das vorgegebene und zum Teil als zu schnell empfundene Tempo der Videopräsentation kritisch angemerkt. Zudem wurden die Videos teilweise sogar als „störend“ wahrgenommen, da diese vom gesprochenen Text ablenkten.

Ein vergleichbares Ergebnis zur Akzeptanz des Verfahrens berichtet Möller (2010) in seiner Dissertation. Hier beurteilten Probanden die soziale Validität des videobasierten Tests „Fragebogen zum Verhalten gegenüber Hotelgästen“ gleicher Höhe wie die textbasierte Version. Mayer et al. (Mayer J. D., Salovey, Caruso, & Sitarenios, 2003) belegten für den MSCEIT, dass eine online dargebotene Version des Tests zu vergleichbaren Ergebnissen führte wie die Papier- und Bleistift-Version. Bei Kanning et al. (Kanning, Grewe, Hollenberg, & Hadouch, 2006) schätzten Probanden nur dann multimediale Items besser ein als textbasierte, wenn ihnen ein allgemeiner Bezugsrahmen in Form einer Übersicht über alle dargebotenen Präsentationsmodi zur Verfügung stand. Dieses Ergebnis lässt sich möglicherweise auf die vorliegende Arbeit übertragen: Die Probanden wurden jeweils direkt nach der Bearbeitung der beiden Testversionen nach ihrer Einschätzung der sozialen Validität gefragt, ohne darüber informiert zu sein, ob bzw. welche Alternativen der Präsentation es grundsätzlich gibt. Vorteil des hier gewählten Vorgehens ist, dass die Einschätzung der Probanden nicht aus einem Vergleich heraus entsteht, sondern aus den Eindrücken der konkreten Testbearbeitung.

Korrelation zu Merkfähigkeit und Aufmerksamkeit. Die beiden Versionen Video und Text unterscheiden sich darin, in welcher Präsentationsform die Informationen dargeboten werden. In der Videoversion sehen die Probanden einen kurzen Film während eine Sprecherin Informationen vorliest. Die Textversion besteht ausschließlich aus dem Text der Sprecherin, der allerdings schriftlich dargeboten wird. In beiden Versionen handelt es sich somit um denselben verbalen Inhalt, den die Probanden aufnehmen müssen; die Videoversion ersetzt nicht den gesprochenen Text, sondern fügt dem verbalen Inhalt sogar noch einen visuellen Inhalt hinzu. Bei der Videoversion müssen sich die Probanden die Informationen zur Situation nach einmaligem Sehen der Videos einprägen; die Textversion ermöglicht den Probanden, bei Bedarf die Situationsbeschreibung erneut im Testheft nachzulesen. Da beim ViTEIP die Bearbeitung im Antwortheft immer erst nach der Videopräsentation der Situation erfolgt, wurde überprüft, inwiefern Merkfähigkeit und Aufmerksamkeit für die Testleistung erforderlich sind. Die Annahmen, dass im Vergleich zur textbasierten Version die Korrelation zwischen dem Gesamtergebnis und Merkfähigkeit

(Hypothese 3c) sowie Aufmerksamkeit (Hypothese 3d) bei der videobasierte Version signifikant höher ausfällt, bestätigten sich nicht. Die Ablehnung dieser beiden Hypothesen ist positiv zu interpretieren, da sich in diesen Annahmen die aus der Literatur abgeleitete Befürchtung manifestierte, für die Bearbeitung einer videobasierten Präsentation könnten andere Fähigkeiten als Emotionale Intelligenz von Bedeutung sein. In Bezug auf die Aufmerksamkeit zeigt sich bei der text- und der videobasierten ViTEIP-Version weder ein signifikanter Zusammenhang zur Leistung im ViTEIP noch ein signifikanter Unterschied in der Korrelation zwischen den beiden Präsentationsmodi. Das Ergebnis zur Merkfähigkeit muss mit gewissen Einschränkungen betrachtet werden: da sich in der Stichprobe ein Deckeneffekt für Merkfähigkeit zeigt, ist die Varianz in dieser Variablen und damit die Differenzierungskraft des Verfahrens eingeschränkt.

5.3.1 Fazit: Präsentationsmodus

Ist ein videobasierter Präsentationsmodus für den ViTEIP-V2 geeignet? Insgesamt bestätigten sich die Vorteile, die verschiedene Autoren (Chan & Schmitt, 1997; Lievens & Sackett, 2006; Richman-Hirsch, Olson-Buchanan, & Drasgow, 2000) in videobasierten Tests sehen, in der vorliegenden Arbeit nicht: verbale Intelligenz war für beide Formate im selben Ausmaß von Bedeutung und es gab keine Unterschiede in Bezug auf die Akzeptanz der Versionen. Aus Sicht der Autorin gilt es daher sorgfältig abzuwägen, ob sich der Aufwand lohnt, den die Entwicklung von videobasierten Verfahren mit sich bringt, wenn sich – wie zumindest in der vorliegenden Studie – keine Vorteile für diesen Präsentationsmodus ergeben. Positiv zu verzeichnen ist, dass sich die Korrelationen zu Merkfähigkeit und Aufmerksamkeit in beiden Versionen nicht signifikant unterschieden. Auf Stichprobenebene stellten sich die Zusammenhänge zwischen Emotionaler Intelligenz und Merkfähigkeit bzw. Aufmerksamkeit vergleichbar dar; in keiner Version wurde eine der genannten Fähigkeiten vermehrt gefordert.

Im Kontext eines videobasierten Einsatzes des ViTEIP-V2 zeigte sich ein signifikanter Zusammenhang zwischen der Testleistung und der Gruppengröße; es ist daher zu empfehlen, keine zu großen Testgruppen gleichzeitig zu untersuchen. Einige Teilnehmer der Untersuchungen gaben im Kommentarfeld des ViTEIP-V2 an, sie hätten sich durch schneller ankreuzende Teilnehmer zu einer höheren Bearbeitungsgeschwindigkeit gedrängt gefühlt. Gleichzeitig gaben andere Probanden an, sie hätten sich in den durch langsamere Teilnehmer entstandenen Pausen gelangweilt. In kleineren Gruppen oder einer PC-basierten Einzeldarbietung mit Kopfhörern könnten diese Probleme umgangen werden. Jeder Testteilnehmer könnte so den ViTEIP-V2 nach seinem eigenen Tempo bearbeiten, ohne von der Bearbeitungsgeschwindigkeit anderer beeinflusst zu werden. Es sollte

dennoch systematisch untersucht werden, ab welcher Gruppengröße und in welcher Form sich die Anzahl der Personen auf die Leistungen der Probanden auswirkt.

In der vorliegenden Arbeit wurden situative Hinweisreize als relevante Informationen für die Schlussfolgerung der Emotionen verstanden (siehe Kapitel 1.5.1). Aus den schriftlichen Kommentaren der Probanden wurde deutlich, dass diese Herangehensweise nur bedingt nachvollziehbar war. Viele Anmerkungen bezogen sich darauf, dass in der Videoversion keine Mimik und Gestik der Probanden zu sehen war. Probanden schienen den Eindruck zu haben, ohne diese Informationen kein angemessenes Urteil über Emotionen geben zu können. Zudem sei zu wenig über die Personen bekannt. Damit stellt sich neben der Form der Präsentation der Items die Frage, welche Informationen Probanden benötigen, um eine Situation adäquat einschätzen zu können. Einige Forscher sehen das Vorliegen sogenannter „social cues“ als wichtige Informationsquelle an, um Emotionen zu verstehen (Ekman & Friesen, 1967). Auch in Konzeptualisierungen von Emotionaler Intelligenz wird nonverbalen Hinweisreizen in Gestik, Mimik, Körperhaltung etc. eine zentrale Rolle zugewiesen und beispielsweise das „Erkennen von Gesichtsausdrücken“ in Tests integriert (Mayer, Caruso, & Salovey, 2000). DeBusk und Austin (2011) fanden hingegen keinen Zusammenhang zwischen Emotionaler Intelligenz und der korrekten Identifikation von Gesichtsausdrücken. Ob soziale Hinweisreize wie Mimik und Gestik im Zusammenhang mit Emotionaler Intelligenz tatsächlich nur eine untergeordnete Rolle spielen, ist noch nicht hinreichend geklärt – Erfahrungen im Zusammenhang mit der Entwicklung von Tests zur Erfassung von Sozialer Intelligenz weisen eher in eine andere Richtung (Baumgarten, persönliche Kommunikation am 14.12.2011). Ebenso könnte es sich um eine komplexere Interaktion zwischen Situationsvariablen und „Social cues“ handeln (Carroll & Russell, 1996; Masuda, et al., 2008). Zudem kann argumentiert werden, dass die Fähigkeit, auch mit einer geringen Anzahl an Informationen korrekte Schlussfolgerungen ziehen zu können, charakteristisch ist für eine hohe Ausprägung an Emotionaler Intelligenz.

5.4 Kritische Diskussion der Validität von ViTEIP-V2 und TEMINT

Die vorliegende Arbeit zielte darauf ab, verschiedene methodische ‚Stellschrauben‘ zu betrachten, die einen Einfluss auf die Validität eines Emotionale-Intelligenz-Tests haben können. In den vorangegangenen Kapiteln wurden die Ergebnisse bezüglich der Scorings, der Antwortformate und des Präsentationsmodus diskutiert und Empfehlungen für zukünftige Testentwicklungen und Testanwendungen ausgesprochen. Im Kontext der methodischen Fragestellungen wurde an verschiedenen Stellen auf nicht-hypothesenkonforme Korrelationsmuster der eingesetzten

Emotionale-Intelligenz-Tests zu anderen Verfahren hingewiesen. Wie diese Validitätsergebnisse insgesamt einzuordnen sind, wird nachfolgend diskutiert. Matthews et al. formulierten *“Currently, we do not know whether: (a) “true” EI³⁶ does not exist and so is impossible to measure, (b) one of the existing tests measures “true” EI and the remainder measure existing constructs and trivia, or (c) several existing tests measure distinct facets that could be interrelated in some more complex psychometric model.”* (Matthews, Emo, Roberts, & Zeidner, 2006, p. 17). Diese etwas ernüchternde Formulierung zum Status von Emotionaler Intelligenz beschreibt leider auch die Befundlage, die sich aus der vorliegenden Arbeit für ViTEIP-V2 und TEMINT ergibt, da für beide Tests eine entscheidende Frage nicht geklärt werden konnte: Was messen der ViTEIP-V2 und TEMINT wirklich?

Die Korrelationsmatrizen in den drei Stichproben waren – unabhängig vom verwendeten Scoring – überwiegend durch nicht-signifikante Korrelationen geprägt. Positiv interpretiert bedeutet dies, dass sich die mit dem ViTEIP-V2 bzw. TEMINT erfassten Konstrukte hinreichend von anderen Konstrukten unterscheiden und somit diskriminante Validität besitzen. Durch die positive Korrelation mit verbaler Intelligenz und den überwiegend nicht signifikanten Korrelationen zu den BIG-5-Persönlichkeitsmerkmalen ist zumindest für den ViTEIP-V2 davon auszugehen, dass eine kognitive Fähigkeit erfasst wird. Die Korrelationsmuster passen damit grundsätzlich zur allgemeinen Befundlage zu Emotionaler Intelligenz, wie sie sich in der Metaanalyse von Van Rooy, Viswesvaran und Pluta (2005) und anderen Untersuchungen darstellt (MacCann & Roberts, 2008; Schulte, Ree, & Carretta, 2004): moderate Korrelationen mit (verbaler) Intelligenz, keine oder geringer Korrelation mit Persönlichkeitsmerkmalen. In Kombination mit den mehrheitlich nicht signifikanten Korrelationen zu weiteren konvergenten Konstrukten (Empathie, selbsteingeschätzter Emotionaler Kompetenz) und den nur unter bestimmten Auswertungsmodalitäten positiven Korrelationen zum TEMINT, ist jedoch unklar, ob mit dem ViTEIP-V2 tatsächlich ein Teilbereich von Emotionaler Intelligenz erfasst werden kann.

Diskriminante Validität. Im Sinne der Abgrenzung von bestehenden Konstrukten wurde erwartet, dass der ViTEIP-V2 und TEMINT diskriminante Validität in Bezug auf Persönlichkeit und Merkfähigkeit zeigen. Die maximal moderate Korrelation zu Merkfähigkeitstests ($r \geq 0.24^{**}$, *Tabelle 9*) entsprach den Erwartungen, da sich darin der kognitive Charakter des Tests abbildet. Die Korrelationen zwischen den Emotionale-Intelligenz-Tests und den Skalen des NEO-FFI fielen erwartungsgemäß überwiegend nicht signifikant aus. Bei der Skala Gewissenhaftigkeit ergaben sich über beide Emotionale-Intelligenz-Tests drei positiv-signifikante Korrelationen zu targetbasierten Scorings mit einer maximalen Höhe von $r \leq 0.20^*$. Die Korrelation zu dieser Skala ist überraschend,

³⁶ EI = emotional intelligence

da, wie unter 1.4 beschrieben, aus einer inhaltlichen Nähe eher Zusammenhänge zu Verträglichkeit oder Offenheit für Neues zu erwarten waren.

Da sich die Korrelation zwischen Gewissenhaftigkeit und Emotionaler Intelligenz nicht aus der konzeptuellen Verwandtschaft erklären lässt, besteht die Möglichkeit, dass Probanden, die sich als gewissenhaft beschreiben, einen Bearbeitungsstil aufweisen, der für TEMINT und ViTEIP-V2 von Vorteil ist. So könnten diese Personen beispielsweise weniger Items unbearbeitet lassen, was sich wiederum positiv auf das Gesamtergebnis auswirkt, da Missings in der Auswertung nicht unberücksichtigt bleiben, sondern mit der Antwortoption „Emotion nicht vorhanden“ ergänzt und entsprechend bewertet werden. Die Korrelation zwischen der Anzahl der Missings und der Skala Gewissenhaftigkeit fällt jedoch nicht signifikant aus; dieser Erklärungsansatz trifft somit nicht zu. Betrachtet man das untere Ende der Skala Gewissenhaftigkeit und die damit einhergehende Beschreibung von Personen mit niedrigen Ausprägungen als unsorgfältig, spontan und ungenau handelnd, erscheint ein Zusammenhang zu einer geringen Ausprägung in Emotionaler Intelligenz nicht völlig unplausibel. Da diese Erklärungsversuche Raum für Interpretationen offen lassen, sei zu weiteren Untersuchungen angeregt. Das vorliegende Ergebnis bestätigt dennoch die generelle Befundlage, dass sich Verfahren zur Erfassung von Emotionaler Intelligenz und Persönlichkeitseigenschaften voneinander unterscheiden, wie es beispielsweise auch Mayer et al. (Mayer J. D., Salovey, Caruso, & Sitarenios, 2003), Bickel und Kollegen (Bickel G., et al., 2009) oder Sharma et al. (Sharma, Gangopadhyay, Austin, & Mandal, 2013) für andere Emotionale-Intelligenz-Tests berichten.

Konvergente Validität. Auf das unerwartete Korrelationsmuster zwischen ViTEIP-V2 und TEMINT bei targetbasierten Scorings soll an dieser Stelle nicht mehr eingegangen werden, da diese Diskussion schon im Kontext der neuen biasfreien Scorings geführt wurde (Kapitel 5.1.3). Hypothesenkonform in Bezug auf die konvergente Validität zeigten sich die Korrelationen des ViTEIP-V2 mit verbaler Intelligenz sowohl für Target-Übereinstimmung mit $r=0.30^{**}$ als auch für die Scorings Target-Treffer ($r=0.26^{**}$), Konsens-Anteil ($r=0.24^{**}$), Konsens-Modal ($r=0.22^{**}$), Profil-Gesamtwert ($r=0.28^{**}$), Detektionsleistung ($r=0.24^{**}$) und emotionale Spezifität ($r=0.31^{**}$). Keine signifikante Korrelation fand sich hingegen zu der zweiten erfassten Intelligenzfacette, der figuralen Intelligenz. Auch bei Mayer et al. (Mayer, Salovey, & Caruso, 2008) oder Ciarrochi et al. (Ciarrochi, Chan, & Caputi, 2000) waren die Korrelationen zu verbaler Intelligenz höher als zu anderen Intelligenzfacetten. Inhaltlich ist es nachvollziehbar, dass Probanden mit einer größeren Fähigkeit verbale Bedeutungen zu erfassen, im ViTEIP-V2 besser abschneiden als diejenigen, die hier eine geringere Fähigkeit aufweisen. Der ViTEIP-V2 enthält (vor allem auch im Vergleich zum TEMINT, der nicht mit verbaler Intelligenz korreliert) einen deutlich größeren Anteil an schriftlichem oder

gesprochenem Text, dessen Verständnis für die Bearbeitung wesentlich ist. Dass die Korrelation zu figuraler Intelligenz nicht signifikant ausfällt, kann auch ein Hinweis darauf sein, dass Emotionale Intelligenz möglicherweise nicht in ein generelles Intelligenzmodell einzuordnen ist, sondern es sich „nur“ um eine verbalgesättigte, kognitive Fähigkeit handelt. Denn wenn es sich bei Emotionaler Intelligenz tatsächlich um eine Form von Intelligenz handelt, sollten im Sinne der positiven Mannigfaltigkeit zumindest geringe positive Korrelationen zu anderen mentalen Fähigkeiten zu finden sein, wie Mayer et al. im Zusammenhang mit den von ihnen formulierten Kriterien für Intelligenz schreiben: *„Emotional intelligence should define a set of abilities that are moderately intercorrelated with one another“* (Mayer, Caruso, & Salovey, 2000, p. 271).

Eine andere Erklärung der vorliegenden Korrelationsmuster lässt sich aus dem Buchkapitel von Horn und Blankson (2012, p. 76) ableiten, in welchem unter der Überschrift *„Foundations for Better Understanding of Cognitive Abilities“* das Cattell-Horn-Carroll-Modell dargestellt wird. An der Spitze des CHC-Intelligenzmodells steht ein Generalfaktor „g“; auf dem darunter liegenden zweiten Stratum platzieren die Autoren verschiedene breit angelegte Fähigkeitskomplexe die aus spezifischen Fähigkeiten des ersten Stratums gebildet werden. Es wurde betrachtet, an welcher Stelle innerhalb dieses Modells die in der vorliegenden Arbeit erfassten Fähigkeiten mit den Bezeichnungen verbale, figurale und emotionale Intelligenz, einzuordnen sind. Die Lösung von figuralen Aufgaben fällt in diesem Modell unter die Sekundärfähigkeiten *„Gf: Abilities of reasoning under novel conditions“* sowie *„Gv: Visualization and spatial orientation abilities“*. Die Primärfähigkeit *„Verbales Verständnis“* zählen sie hingegen zu demselben Bereich *„Gc: Acculturational knowledge abilities“*, zu dem auch *„verhaltensbezogene Beziehungen“* gehört. Letztgenanntes beschreiben sie als Beurteilung von Interaktionen zwischen Personen, mit der Einschätzung darüber, wie sich jemand in dieser Situation fühlt. Die Fähigkeit zur Einschätzung von Interaktionen (was ein wesentlicher Bestandteil Emotionaler Intelligenz ist) liegt somit in dieser Beschreibung von Intelligenz konzeptionell deutlich näher bei verbalen Fähigkeiten als bei figuralen Fähigkeiten.

Ein weiterer Ansatz zur Erklärung für geringe/fehlende Korrelationen findet sich bei Kang, Day und Meara (2006). Die Autoren schlagen vor, dass eine Unterscheidung in fluide und kristalline³⁷ Aspekte auch für das Konstrukt Emotionale Intelligenz sinnvoll sein könnte. Konzeptionell betrachtet würde der in TEMINT und ViTEIP-V2 erfasste Teilbereich eher einem wissens- und erfahrungsbasierten (kristallinen) Intelligenzbereich zugeordnet. Die erfassten Facetten der

³⁷ Unter kristalliner Intelligenz wird in der Psychologie die Fähigkeit verstanden, kulturabhängiges Wissen zu erwerben, wiederzugeben und anzuwenden. Fluide Intelligenz bezieht sich auf die Fähigkeit zu logischem, induktivem und deduktivem Denken.

verbalen und figuralen Intelligenz werden jedoch beide zum sogenannten „Grundmodul“ des IST-2000-R gezählt, das Kennwerte für fluide Intelligenz liefert. Würden kristalline Intelligenzfacetten erhoben, fiel die Korrelationen zwischen Emotionaler Intelligenz und Intelligenz vermutlich insgesamt deutlich höher aus.

Die fehlenden signifikanten Korrelationen des TEMINT zu Intelligenz überraschten, da es verschiedene Studien gibt, aus denen die Validität des TEMINT als gut belegt zu interpretieren ist (Blickle G. , et al., 2009; Blickle G. , Momm, Liu, Witzki, & Steinmayr, 2011). Doch gleichzeitig existieren für den TEMINT ebenso Studien mit weniger deutlichen Zusammenhängen, wie beispielsweise bei Blickle et al. (Blickle, Kramer, & Mierke, 2010) mit nicht signifikanten Korrelationen zu Intelligenz. Auch die Korrelation mit anderen Emotionale-Intelligenz-Maßen fällt für den TEMINT teilweise nicht signifikant aus (Choi & Kluemper, 2012). Da in der vorliegenden Arbeit vor allem die Items mit der Angabe der Originalperson „Emotion nicht vorhanden“ von vielen bzw. zum Teil von allen Probanden korrekt beantwortet wurden (*Tabelle A-4 in Anhang C*), könnte die Schwierigkeit des TEMINT insgesamt zu gering sein, was wiederum zu Deckeneffekten mit geringeren Differenzierung zwischen verschiedenen Leistungsniveaus führt. Dies zu untersuchen und gegebenenfalls den TEMINT um schwierigere Items zu ergänzen, soll hiermit angeregt werden.

Für die Überprüfung der konvergenten Validität des ViTEIP-V2 und TEMINT wurden zudem die Zusammenhänge zum Emotionale-Kompetenz-Fragebogen EKF (alle drei Stichproben) und zum situationsbezogenen Empathiefragebogen (Stichproben 2a und 2b) betrachtet. Die Korrelationen mit dem ViTEIP-V2 bzw. TEMINT fielen bei fast allen Scorings für beide Tests über die Stichproben hinweg überwiegend nicht signifikant aus. Werden Empathie und Emotionale Kompetenz im Sinne eines Trait-Ansatzes verstanden, ist eine geringe oder gar fehlende Korrelation zu Fähigkeits-Tests wie dem ViTEIP-V2 oder TEMINT nicht überraschend. Hier ist die Befundlage mittlerweile relativ eindeutig und besagt, dass es zwischen Trait- und Fähigkeits-Emotionaler-Intelligenz nur einen sehr geringen Überschneidungsbereich gibt (Christiansen, Janovics, & Siers, 2010; Zeidner, Matthews, & Roberts, 2001). Dies resümierten auch Joseph und Newman (2010) aus ihrer Metaanalyse, in der sie verschiedene Emotionale-Intelligenz-Ansätze und -Methoden untersuchten und eine mittlere Korrelation zwischen Trait- und Fähigkeits-Emotionaler-Intelligenz von $r=0.14$ berechneten. Auch in anderen Untersuchungen korrelierte Emotionale Intelligenz nicht mit Empathie (z.B. Geher, Warner, & Brown, 2001), was möglicherweise auch auf die theoretische Unschärfe des Konstrukts Empathie sowie die zum Teil geringe Testgüte der Messinstrumente zurückzuführen ist. Dies kann für den in der vorliegenden Arbeit eingesetzten Empathietest nicht ausgeschlossen werden, der bislang nur in wenigen Untersuchungen zum Einsatz kam.

Es scheint sich die Vermutung zu bestätigen, dass es bei Emotionaler Intelligenz/Kompetenz für die Probanden schwierig ist, das eigene Fähigkeitsniveau korrekt einzuschätzen (vgl. Kapitel 1.2). Die wahrgenommene Emotionale Intelligenz durch die Person selbst kann sich nach Ciarrochi et al. (Ciarrochi, Chan, Caputi, & Roberts, 2001) deutlich von dem in einem Leistungstest erfassten Niveau unterscheiden. In Stichprobe 1, die sich überwiegend aus Psychologiestudenten zusammensetzt, zeigten sich einige signifikant-positive Zusammenhänge zu EKF-Facetten. Dies deutet darauf hin, dass hier die Selbsteinschätzungen und die tatsächlichen Leistungen bezüglich emotionaler Fähigkeiten eher kongruent ausfallen als in den anderen Stichproben. Warum dies nicht auf den TEMINT zutrifft, ist unklar. Möglicherweise spielt auch hier die geringe Schwierigkeit des TEMINT eine Rolle, die zu einer geringen Differenzierungskraft des Tests führt.

Neben einem fehlenden Zugang zum eigenen Ausprägungsgrad der emotionalen Fähigkeiten bzw. Empathie, können auch Verzerrungen aufgrund eines sozial erwünschten oder selbstwertdienlichen Antwortverhaltens in den Selbsteinschätzungen auftreten. Sowohl für den EKF als auch den Empathietest sind aufgrund des Selbstbeschreibungsformats sozial erwünschte Antworten der Probanden möglich. Gerade bei Personen, die einen Beruf anstreben, in dem sie mit Menschen arbeiten und auf deren Befindlichkeiten sie eingehen müssen, kann eine bewusste oder unbewusste Neigung vorliegen, sich im Bereich der emotionalen oder empathischen Kompetenzen gut darzustellen. Diese Neigung muss jedoch nicht automatisch einer sozial erwünschten Komponente entsprechen, sondern kann auch im Sinne einer selbstwertdienlichen Selbsteinschätzung verstanden werden. Würden Personen ihre eigenen emotionalen Fähigkeiten als gering erachten, müssten sie damit auch ihre Eignung für den gewählten Beruf als Psychologe oder Pflegekraft in Frage stellen, was kognitive Dissonanz auslösen könnte. Die ursprüngliche Versuchsplanung, über die EKF-Fremdeinschätzung Ergebnisse zu erhalten, die der Selbsteinschätzung gegenübergestellt werden können, konnte leider aufgrund eines zu geringen Rücklaufs der Fremdeinschätzungen nicht realisiert werden.

Auch die spezifische Aufgabenstellung der Tests kann zu Unterschieden in den Ergebnissen führen (vgl. Ployhart & Ehrhart, 2003). Im ViTEIP-V2 geht es darum, anhand einer Situationsbeschreibung die Intensität der Emotionen der Originalperson zu schlussfolgern. Im Empathietest werden zwar auch Situationsbeschreibungen dargeboten, die Aufgabe der Probanden lautet jedoch „Wie stark haben Sie am Ende der Schilderung den Eindruck zu wissen, was in der anderen Person gerade vor sich geht?“. Ein Proband könnte sich somit nach eigener Einschätzung sicher sein zu wissen, wie es um die Gefühlslage der Originalperson aussieht, dennoch aber eine inkorrekte Bewertung über die tatsächlichen Emotionen vornehmen – was zu einem hohen Testwert im Empathietest und einem geringen Testwert im ViTEIP führen würde.

Fazit. Der aufmerksame Leser mag sich bei den Ausführungen an das Kapitel 1.4 „Die Validität von Emotionale-Intelligenz-Tests“ zu Beginn dieser Arbeit erinnern. Dort wurde ausgeführt, dass die Validierung von Tests zur Erfassung von Emotionaler Intelligenz insgesamt als noch nicht ausreichend betrachtet werden kann und somit nicht nur ein Spezifikum dieser Dissertation bzw. des ViTEIP-V2 und TEMINT darstellt. Die kritische Diskussion der vorliegenden Validitätsergebnisse soll daher in erster Linie dazu anregen, den neu entwickelten ViTEIP-V2 in weiteren Untersuchungen in Ergänzung zu etablierten Verfahren einzusetzen. Gerade unter Berücksichtigung der Ergebnisse, die sich bei Anwendung der ‚neuen‘ Scorings ergeben – und hier vor allem bei der getrennten Auswertung von Emotionaler Sensitivität und Emotionaler Spezifität – erscheint es sinnvoll und vielversprechend, die unterschiedlichen Validitätsansprüche, die diese Kennwerte zeigen, empirisch weiter abzusichern. Gleiches gilt für den TEMINT, für den sich in der vorliegenden Arbeit unerwartet andere Korrelationsmuster zeigten als in vorangegangenen Untersuchungen.

5.5 Anregungen für zukünftige Studien

In den vorliegenden Untersuchungen konnten einige Aspekte nicht systematisch genug untersucht werden und es sind Fragen offen geblieben bzw. entstanden. Im Sinne von Anregungen für zukünftige Untersuchungen und Testentwicklungen im Bereich Emotionale Intelligenz soll auf diese Punkte eingegangen werden.

Untersuchungen zu Kriteriumsvalidität. Matthews, Zeidner und Roberts (2012) kritisierten an der aktuellen Forschung zu Emotionaler Intelligenz, dass nach wie vor schlüssige Belege für die Kriteriumsvalidität fehlen. In der zweiten Untersuchung sollten daher anhand von Einschätzungen durch Kollegen und Vorgesetzte verschiedene externe Kriterien und der EKF in der Fremdeinschätzung erhoben werden. Trotz gründlicher Vorbereitung und konsequentem Nachhalten war der Rücklauf der Fremdbeurteilungen so gering, dass Berechnungen zwischen der Leistung im ViTEIP-V2 bzw. TEMINT und der Eignungseinschätzung bzw. der Fremdeinschätzung der Emotionalen Kompetenz nicht sinnvoll waren. Die Autorin schließt sich dem Appell von Webb et al. (Webb, et al., 2013) an, sich dieser Fragestellung zukünftig intensiver zu widmen, um damit auch den tatsächlichen, praktischen Nutzen von Emotionaler Intelligenz zu verdeutlichen. Erst wenn die praktische Relevanz hinreichend belegt werden kann, liegt in einem Konstrukt Emotionale Intelligenz ein echter Gewinn für Forschung und Praxis.

Untersuchungen zur Struktur von Emotionaler Intelligenz. Grundsätzlich stellt sich die Frage, ob für die Konzeption von Emotionaler Intelligenz andere Unterscheidungsmerkmale oder Taxonomien erforderlich sind als die bisher verwendeten. Mayer et al. (Mayer, Caruso, & Salovey, 2000)

unterscheiden in ihrem Modell (und in ihrem Test MSCEIT) vier verschiedene Fähigkeiten, die in der Komplexität und Schwierigkeit aufeinander aufbauen. Zeidner et al. (Zeidner M., Matthews, Roberts, & MacCann, 2003) hingegen gehen in ihrem Multi-Level Investment Model davon aus, dass untergeordnete Kompetenzen die Basis für die Entwicklung höherer Kompetenzen wie die Emotionsregulation bilden. In der vorliegenden Arbeit zeigte sich, dass die Unterscheidung im Sinne der Signalentdeckung (getrennte Betrachtung von „Signal“ und „Rauschen“) die Ergebnisse besser erklärte als unter Verwendung anderer Scorings. Auch wenn die SET zunächst als Methodik für ein biasfreies Scoring herangezogen wurde, könnte darin auch mehr als nur ein Auswertungsansatz für Emotionale Intelligenz liegen. Durch die Übertragung der Grundgedanken der Signalentdeckungstheorie auf Emotionale Intelligenz ergeben sich neue Annahmen und Interpretationen, die der Diskussion um das Konstrukt eine neue Wendung geben können.

Bei den explorativen faktorenanalytischen Untersuchungen des ViTEIP und TEMINT konnte die von Mayer et al. (2003) postulierte Annahme eines Emotionale-Intelligenz-Generalfaktors nicht bestätigt werden. Ein Ergebnis der vorliegenden Arbeit ist, Emotionale Intelligenz nicht als ein homogenes Konstrukt zu verstehen, sondern in zwei Komponenten „Emotionale Sensitivität“ und „Emotionale Spezifität“ zu unterteilen. Auch Baumgarten (persönliche Kommunikation am 14.12.2011) berichtete von ähnlichen Ergebnissen bei der Testentwicklung im Bereich Soziale Intelligenz: uneinheitliche Korrelationen zwischen Target- und Konsensscorings, negative Trennschärfen innerhalb von Soziale-Intelligenz-Skalen, geringe interne Konsistenzen, heterogene Ergebnisse über Stichproben hinweg, geringe oder Nullkorrelationen bei Validitätsuntersuchungen (z.B. zu Intelligenz, Persönlichkeit, Emotionale Intelligenz). Gerade weil es sich bei Sozialer und Emotionaler Intelligenz um zwei distinkte Konstrukte handelt, ist es interessant, dass die Magdeburger Arbeitsgruppe nach Baumgarten bei Sozialer Intelligenz nicht mehr von einem einheitlichen Konstrukt ausgeht, sondern von sozialer Wahrnehmung, sozialem Verständnis, sozialem Wissen etc. sprechen, da diese Facetten – anders als ursprünglich postuliert – untereinander nur gering korrelieren (Conzelmann, Weis, & Süß, 2013).

Die Ergebnisse der vorliegenden Dissertation stellen die weit verbreitete Annahme, bei Emotionaler Intelligenz handele es sich um ein eindimensionales und homogenes Konstrukt in Frage und regen zu weiteren Forschungen bezüglich der Struktur dieses Konstrukts an.

Untersuchung verschiedener Scorings in Verbindung mit anderen methodischen Aspekten. In der vorliegenden Arbeit zeigte sich, dass die Auswertungsstrategien in erster Linie in Kombination mit anderen methodischen Aspekten eine Auswirkung auf Testergebnisse haben. Es ergab sich, dass die Leistungen im ViTEIP-V2 grundsätzlich nicht vom Präsentationsmodus (Video vs. Text) abhängen,

sich die Ergebnisse jedoch in Kombination mit dem Scoring verändern. Probanden schnitten bei targetbasierten Auswertungen in der Textversion, bei konsensbasierten Auswertungen bei der Videoversion besser ab. Ein ähnliches Bild ergab sich auch bei der Analyse der Auswirkungen des Antwortformats. Werden targetbasierte Auswertungen verwendet, fielen die Varianz und die Reliabilität bei „Ankreuzen“ größer als bei „Zuordnen“ aus. Bei konsensbasierten Scorings waren Varianz und Reliabilität bei „Zuordnen“ größer. Die Vergleichbarkeit der Auswertungsprozeduren scheint also nicht unter allen Bedingungen gleichermaßen gegeben zu sein; andere methodische Aspekte müssen zusätzlich berücksichtigt werden. Um zu untersuchen, wie das genaue Zusammenspiel von Scoring, Präsentationsmodus und Antwortformat aussieht, bietet es sich an, zukünftige Untersuchungen so zu konzipieren, dass alle drei Faktoren systematisch analysiert werden können. Aus den Daten der vorliegenden Arbeit ist eine solch differenzierte Analyse nicht möglich.

5.6 Limitationen

In jeder empirischen Arbeit liegen Einschränkungen in der Übertragbarkeit und Interpretierbarkeit der Ergebnisse vor, die sich aufgrund der Rahmenbedingungen ergeben.

Einige Limitationen beziehen sich auf den ViTEIP-V2, der als neues Verfahren für die Erfassung von Emotionaler Intelligenz in dieser Arbeit erstmals eingesetzt wurde. Auf die Stichprobenszusammensetzung der Originalpersonen (ausschließlich Frauen) sowie das Überwiegen von negativen Emotionen wurde in Kapitel 1.5.3 schon kurz verwiesen, soll hier jedoch vertieft werden. Beide Aspekte ergeben sich aus dem Konstruktionsprinzip des ViTEIP, bei dem Bewohner einer Altenhilfeeinrichtung und Patienten einer Rehaklinik gebeten wurden, tatsächlich erlebte Situationen zu schildern. Diesem Aufruf folgten in erster Linie Frauen und es wurden insgesamt vor allem Situationen benannt, in denen negative Emotionen im Mittelpunkt standen; auf dieser Grundlage wurden die Situationen und Emotionen der Endversion des ViTEIP generiert. Außer Acht bleibt hier, dass in Einrichtungen des Gesundheitswesens selbstverständlich auch männliche Bewohner/Patienten betreut werden und Pflegekräfte mit den Emotionen dieser Gruppe professionell umzugehen haben. Inwiefern sich Situationen, die von Frauen oder Männern geschildert werden, voneinander unterscheiden oder ob es Probanden unterschiedlich gut gelingt, die vorhandenen Emotionen zu identifizieren, kann aus der vorliegenden Untersuchung nicht gefolgert werden. Der ViTEIP-V2 bildet somit nur einen Teil des Pflegealltags ab. Ob das Geschlecht der Originalpersonen eine Rolle für die Probanden spielt, konnte nicht überprüft werden, da im ViTEIP-V2 ausschließlich Situationen mit weiblichen Originalpersonen vorkommen. Hier könnte eine

Untersuchung Aufschluss bringen, in der den Probanden beispielsweise dieselbe Situation einmal mit einer weiblichen und einmal mit einer männlichen Originalperson dargeboten werden.

Eine weitere Limitation des ViTEIP-V2 liegt darin, dass in den von den Originalpersonen geschilderten Situationen überwiegend negative Emotionen vorherrschen. Damit deckt der ViTEIP nur einen Teilbereich von „Erkennen von Emotionen Anderer anhand situativer Auslöser“ ab; der Umgang mit positiven Emotionen bleibt weitestgehend unberücksichtigt. Dem entgegensustellen ist, dass im Alltag einer Pflegekraft vor allem der Umgang mit Situationen, die patientenseitig mit negativen Emotionen einhergehen, als die schwierigeren empfunden werden und weniger die Situationen, die positive Emotionen wie Freude, Stolz oder Zuneigung auslösen. Möglicherweise sind es die Begebenheiten mit negativen Emotionen, die Patienten vermehrt in Erinnerung bleiben und somit in den Schilderungen überwiegen. Der ViTEIP-V2 ist somit ökologisch valide; strenggenommen können jedoch in erster Linie Aussagen gemacht werden über die Fähigkeit, negative Emotionen zu identifizieren.

Ein wesentlicher Kritikpunkt ist in der vorliegenden Arbeit, dass neben dem neu entwickelten ViTEIP-V2 mit dem TEMINT nur ein weiterer Emotionale-Intelligenz-Leistungstest eingesetzt wurde. Da sich beim TEMINT zu anderen Verfahren überwiegend Null-Korrelationen zeigten, ist unklar, welches Konstrukt mit diesem Test erfasst wird. Sich für die Validierung eines neuen Verfahrens nur auf ein weiteres Instrument zum selben Konstrukt zu verlassen, erscheint im Rückblick kritisch, zumal in der zweiten Datenerhebung der Rücklauf der Fremdeinschätzungen sehr viel geringer ausfiel als erwartet, so dass weder externe Kriterien noch EKF-Fremdeinschätzungen in die Analysen mit aufgenommen werden konnten. Aufgrund der allgemein bekannten Schwierigkeiten mit Rückläufen von Fragebögen, wurde in der Vorbereitung der Studie zwar viel Zeit investiert, um die Chancen auf eine angemessene Anzahl an Rückmeldungen zu erhöhen. Trotz dieser sorgfältigen Planungen und aller Bemühungen im Nachgang, an weitere Fremdeinschätzungen zu kommen, konnten keine ausreichend großen Rückläufe erzielt werden, die sinnvolle Analysen ermöglichen. Damit bleiben verschiedene Fragen der (Kriteriums-)Validität ungeklärt.

Nachdem sich in Untersuchung 1 die vermuteten Zusammenhänge zwischen Emotionaler Intelligenz, Intelligenz und Persönlichkeitsmerkmalen bestätigten, wurde in Untersuchung 2 auf eine entsprechende Analyse verzichtet. Rückblickend kann dies als Limitation gesehen werden, da es aufgrund der generellen Heterogenität der Ergebnisse von Interesse wäre, wie sich in dieser Stichprobe der Zusammenhang darstellt. Gegebenenfalls ergäben sich durch Kenntnis der Zusammenhänge neue/andere Interpretationsmöglichkeiten der vorliegenden Ergebnisse.

Eine Einschränkung, die sich aus dem Studiendesign ergibt liegt darin, dass keine Aussagen über Re-Test-Reliabilitäten des ViTEIP-V2 und TEMINT gemacht werden können. Aus organisatorischen und ökonomischen Gründen war es nicht möglich, derselben Stichprobe nach einer gewissen Zeit die Tests ein zweites Mal zur Bearbeitung vorzulegen. Aus den eben genannten Gründen war es auch nicht umzusetzen, die Probanden der Stichprobe 2 den ViTEIP-V2 sowohl mit dem Antwortformat „Ankreuzen“ als auch „Zuordnen“ bearbeiten zu lassen. Die in der vorliegenden Arbeit durchgeführten Vergleiche beziehen sich damit immer auf die Teilstichproben 2a und 2b und nicht auf intraindividuelle Unterschiede. Als ergänzende Erklärung muss an dieser Stelle hinzugefügt werden, dass die ursprüngliche Zielrichtung der Arbeit nicht auf der Untersuchung der methodischen Fragen, sondern auf der Validierung des ViTEIP-V2 lag. Die aktuelle Schwerpunktsetzung kristallisierte sich erst nach der Durchführung der beiden Untersuchungen heraus. Daher wurden die Untersuchungen unter etwas anderen Vorzeichen konzipiert, als man es idealerweise zur systematischen Analyse der methodischen Fragen getan hätte. Zu Untersuchungen, die Aussagen über Re-Test-Reliabilitäten sowie einen direkten Vergleich der beiden Antwortformate ermöglichen, soll an dieser Stelle explizit angeregt werden.

6 Anhang

Übersicht

Anhang A.	<i>ViTEIP-V2</i>	124
Anhang B.	<i>Profilauswertung</i>	126
Anhang C.	<i>Prozentuale Verteilung der Probandenantworten auf die Antwortoptionen</i>	128
Anhang D.	<i>Deskriptive Statistiken</i>	132
Anhang E.	<i>Faktorenanalysen ViTEIP-V2 und TEMINT</i>	133
Anhang F.	<i>Formelsammlung</i>	142
Anhang G.	<i>Antwortbias auf Seiten der Probanden</i>	144
Anhang H.	<i>Übertragung der Signalentdeckungstheorie auf Emotionale-Intelligenz-Tests</i>	146
Anhang I.	<i>Vergleich der Antwortformate</i>	150
Anhang J.	<i>Analyse der Präsentationsmodi</i>	152
Anhang K.	<i>Die Geschichte mit dem Hammer (Watzlawick, 2004)</i>	157

Tabellenverzeichnis Anhang

Tabelle A-1:	<i>Überblick über die Situationen des ViTEIP-V2</i>	124
Tabelle A-2:	<i>Korrelation „Profil-Gesamtwert“ ViTEIP-V2 und TEMINT mit anderen Verfahren</i>	127
Tabelle A-3:	<i>Angaben der Originalpersonen im Vergleich zur Häufigkeit der von den Probanden gewählten Antwortoptionen (ViTEIP-V2)</i>	128
Tabelle A-4:	<i>Angaben der Originalpersonen im Vergleich zur Häufigkeit der von den Probanden gewählten Antwortoptionen (TEMINT)</i>	130
Tabelle A-5:	<i>Deskriptive Statistiken der eingesetzten Testverfahren</i>	132
Tabelle A-6:	<i>Strukturmatrix einer Hauptkomponentenanalyse der Situationen des ViTEIP-V2, Stichprobe 1, Target-Übereinstimmung</i>	133
Tabelle A-7:	<i>Komponentenkorrelationsmatrix</i>	133
Tabelle A-8:	<i>Strukturmatrix einer Hauptkomponentenanalyse der Situationen des ViTEIP-V2, Stichprobe 1, Target-Treffer</i>	134
Tabelle A-9:	<i>Komponentenkorrelationsmatrix</i>	134
Tabelle A-10:	<i>Strukturmatrix einer Hauptkomponentenanalyse der Situationen des ViTEIP-V2, Stichprobe 1, Konsens-Anteil</i>	135
Tabelle A-11:	<i>Komponentenkorrelationsmatrix</i>	135

<i>Tabelle A-12: Strukturmatrix einer Hauptkomponentenanalyse der Situationen des ViTEIP-V2, Stichprobe 2a, Target-Übereinstimmung</i>	<i>136</i>
<i>Tabelle A-13: Komponentenkorrelationsmatrix.....</i>	<i>136</i>
<i>Tabelle A-14: Strukturmatrix einer Hauptkomponentenanalyse der Situationen des TEMINT, Stichprobe 1, Target-Übereinstimmung</i>	<i>137</i>
<i>Tabelle A-15: Komponentenkorrelationsmatrix.....</i>	<i>137</i>
<i>Tabelle A-16: Strukturmatrix einer Hauptkomponentenanalyse der Situationen des TEMINT, Stichprobe 1, Target-Treffer</i>	<i>138</i>
<i>Tabelle A-17: Komponentenkorrelationsmatrix.....</i>	<i>138</i>
<i>Tabelle A-18: Strukturmatrix einer Hauptkomponentenanalyse der Situationen des TEMINT, Stichprobe 1, Konsens-Anteil</i>	<i>139</i>
<i>Tabelle A-19: Komponentenkorrelationsmatrix.....</i>	<i>139</i>
<i>Tabelle A-20: Strukturmatrix einer Hauptkomponentenanalyse der Situationen des TEMINT, Stichprobe 2a, Target-Übereinstimmung</i>	<i>140</i>
<i>Tabelle A-21: Komponentenkorrelationsmatrix.....</i>	<i>140</i>
<i>Tabelle A-22: Strukturmatrix einer Hauptkomponentenanalyse der Situationen des TEMINT, Stichprobe 2a, Target-Treffer</i>	<i>141</i>
<i>Tabelle A-23: Komponentenkorrelationsmatrix.....</i>	<i>141</i>
<i>Tabelle A-24: Korrelation der Kennwerte zum Antwortverhalten mit den Gesamtestwerten bei unterschiedlichen Scorings (ViTEIP-V2 bzw. TEMINT).....</i>	<i>144</i>
<i>Tabelle A-25: Korrelation der Häufigkeiten der gewählten Antwortoptionen bei ViTEIP-V2 und TEMINT</i>	<i>145</i>
<i>Tabelle A-26: Korrelation des ViTEIP-V2 und TEMINT mit anderen Verfahren (Stichprobe 2a)</i>	<i>146</i>
<i>Tabelle A-27: Korrelation des ViTEIP-V2 und TEMINT mit anderen Verfahren (Stichprobe 2b)</i>	<i>146</i>
<i>Tabelle A-28: Korrelationsmatrix SET-Kennwerte des ViTEIP-V2 und TEMINT (Stichprobe 1).....</i>	<i>147</i>
<i>Tabelle A-29: Korrelationsmatrix SET-Kennwerte des ViTEIP-V2 und TEMINT (Stichprobe 2a).....</i>	<i>148</i>
<i>Tabelle A-30: Korrelationsmatrix SET-Kennwerte des ViTEIP-V2 und TEMINT (Stichprobe 2b).....</i>	<i>149</i>
<i>Tabelle A-31: Akzeptanzwerte des ViTEIP-V2 bei Antwortformat „Ankreuzen“ und „Zuordnen“ ...</i>	<i>150</i>
<i>Tabelle A-32: Korrelation der Kennwerte des ViTEIP-V2 bei unterschiedlichen Antwortformaten und verschiedenen Scorings zu anderen Variablen</i>	<i>151</i>
<i>Tabelle A-33: Mittelwertvergleich der text- und der videobasierten Forschungsversion (ViTEIP; Scoring: Target-Übereinstimmung)</i>	<i>152</i>
<i>Tabelle A-34: Mittelwertvergleich der text- und der videobasierten Forschungsversion (ViTEIP; Scoring: Target-Treffer)</i>	<i>153</i>
<i>Tabelle A-35: Mittelwertvergleich der text- und der videobasierten Forschungsversion (ViTEIP; Scoring: Konsens-Anteil)</i>	<i>154</i>

<i>Tabelle A-36: Mittelwertvergleich der text- und der videobasierten Forschungsversion (ViTEIP; Scoring: Konsens-Modal)</i>	<i>155</i>
<i>Tabelle A-37: Deskriptive Statistiken von „Text“ und „Video“ bei unterschiedlichen Scorings (ViTEIP).....</i>	<i>155</i>
<i>Tabelle A-38: Vergleich der Akzeptanzwerte der text- und der videobasierten Forschungsversion (ViTEIP).....</i>	<i>156</i>
<i>Tabelle A-39: Korrelationsmatrix ViTEIP „Text“ und „Video“ bei unterschiedlichen Scorings</i>	<i>156</i>

Abbildungsverzeichnis Anhang

<i>Abbildung A-1: Beispiel für ViTEIP-V2 Antwortformat „Ankreuzen“.....</i>	<i>125</i>
<i>Abbildung A-2: Beispiel für ViTEIP-V2 Antwortformat „Zuordnen“.....</i>	<i>125</i>
<i>Abbildung A-3: Visualisiertes Beispiel für ViTEIP-V2 Profilauswertung.....</i>	<i>126</i>

Anhang A. ViTEIP-V2*Tabelle A-1: Überblick über die Situationen des ViTEIP-V2*

Nr.	Patient	Ereignis	Dominante Emotionen ³	Item- anzahl	„Null- Items“
1	Frau Klein ¹	Zu Bett gebracht werden	Abneigung, Ärger , Trauer, Schuld	8	3
2	Frau Schmidt ¹	Ein Glas zerbricht	Dankbarkeit , Scham, Freude, Zuneigung	11	4
4	Frau Fischer ¹	Abgebrochener Besuch	Ärger, Trauer, Enttäuschung , Abneigung	8	3
5	Frau Schubert ²	Unerwartete Nachvisite	Überraschung , Abneigung, Ärger	6	2
6	Frau Albrecht ²	Milchunverträglichkeit	Trauer, Überraschung , Enttäuschung	6	3
7	Frau Hoffmann ¹	Bestellung	Angst, Unruhe , Überraschung	9	5
9	Frau Schubert ²	Plötzlicher Durchfall	Scham , Abneigung, Enttäuschung, Ärger, Ekel, Schuld, Überraschung	11	3
11	Frau Fischer ¹	Wellensittich gefunden	Freude, Stolz , Dankbarkeit	6	1
12	Frau Krüger ²	Sturz ohne Korsett	Angst	10	5
13	Frau Schubert ²	Ausbleibender Anruf	Trauer, Enttäuschung , Ärger	9	3
15	Frau Ludwig ²	Plötzliches Weinen	Angst, Unruhe, Trauer	7	4
16	Frau Fischer ¹	Bitte um Toilettenpapier	Abneigung, Ärger	8	4
Gesamttest				99	40

Anmerkungen:

Nummerierung der Situationen entspricht den Nummerierungen der ViTEIP-Version V1.

„Null-Items“= Items, bei denen die Antwort der Originalperson „Emotion nicht vorhanden“ lautet.

¹ Die Person ist Bewohnerin der Altenhilfeeinrichtung.

² Die Person ist Patientin in der Rehaklinik.

³ **fett** gedruckte Emotionen wurden von den Originalpersonen als „stark vorhanden“, normal gedruckte als „mittel vorhanden“ angegeben; Emotionen, die „schwach vorhanden“ oder „nicht vorhanden“ sind, werden hier nicht aufgeführt.

Situation 4: Frau Fischers Familie kommt zu Besuch				
Wie stark waren Frau Fischers Gefühle zum Ende dieser Situation?				
	nicht vorhanden	schwach vorhanden	mittel vorhanden	stark vorhanden
Abneigung	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Ärger	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Neid	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Angst/Furcht	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Unruhe	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Traurigkeit	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Enttäuschung	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Schuld (-gefühl)	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Scham	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Freude	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Stolz	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Dankbarkeit	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Zuneigung	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Überraschung	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Ekel	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Abbildung A-1: Beispiel für ViTEIP-V2 Antwortformat „Ankreuzen“

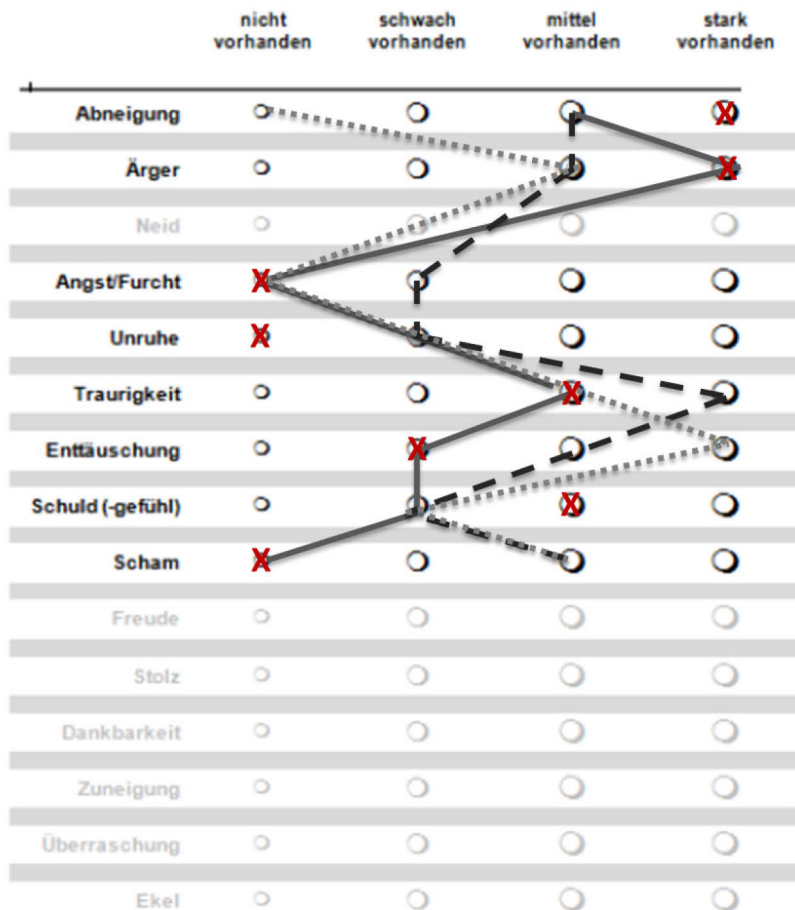
Situation 4: Frau Fischers Familie kommt zu Besuch	
Wie stark waren Frau Fischers Gefühle zum Ende dieser Situation?	
Bitte ordnen Sie die untenstehenden Emotionen in die Ihrer Meinung nach passenden Gruppen (Kästen) ein und streichen Sie diese im unteren Kasten durch:	
<div>Stark vorhandene Emotionen</div> <div></div>	
<div>Mittel vorhandene Emotionen</div> <div></div>	
<div>Schwach vorhandene Emotionen</div> <div></div>	
Bitte ordnen Sie folgende Emotionen zu:	
<div> <div>Situationen, die in diesem Kasten verbleiben werden als nicht vorhanden eingestuft</div> <div> <div>Abneigung</div> <div>Ärger</div> <div>Angst/Furcht</div> </div> <div> <div>Traurigkeit</div> <div>Enttäuschung</div> <div>Schuld (-gefühl)</div> <div>Scham</div> </div> <div>Überraschung</div> </div>	

Abbildung A-2: Beispiel für ViTEIP-V2 Antwortformat „Zuordnen“

Anhang B. Profilauswertung

Situation 1: Frau Klein wird zu Bett gebracht

Wie stark waren Frau Kleins Gefühle zum Ende dieser Situation?



X = AdO

VP 74 $r_{ss} = 0.89$

VP 14 $r_{ss} = 0.51$

VP 27 $r_{ss} = 0.08$

Anmerkungen:

AdO= Angabe der Originalperson, d.h. welche Ausprägung der Emotion hat die Originalperson angegeben?

VP= Versuchsperson, hier exemplarisch für drei Versuchspersonen.

r_{ss} = Spearman Rangkorrelation, siehe Kapitel 2.1.1.1

Abbildung A-3: Visualisiertes Beispiel für ViTEIP-V2 Profilauswertung

Tabelle A-2: Korrelation „Profil-Gesamtwert“ ViTEIP-V2 und TEMINT mit anderen Verfahren

	Profil-Gesamtwert ViTEIP-V2			Profil-Gesamtwert TEMINT		
	Stichprobe 1	Stichprobe 2a	Stichprobe 2b	Stichprobe 1	Stichprobe 2a	Stichprobe 2b
d2-R_ Konzentrationsleistung	-0.10	0.07	0.02	-0.10	0.07	0.08
IST_Verbal	0.28**			-0.08		
IST_Figural	0.08			-0.02		
IST_Merkfähigkeit bzw. WIT	0.18*	-0.03	0.05	-0.02	0.17*	0.09
Empathiefragebogen		-0.10	0.08		0.03	0.01
EKF_Gesamt	0.13	-0.11	-0.06	0.02	-0.16*	-0.10
EKF_EE	0.16*	-0.01	0.04	0.01	-0.13	-0.11
EKF_EA	0.09	-0.11	-0.07	0.08	-0.15*	-0.01
EKF_RG	0.13	-0.04	-0.10	-0.02	-0.16*	-0.07
EKF_EX	0.03	-0.16*	-0.07	0.00	-0.01	-0.10
NEO-FFI_Neurotizismus	-0.07			-0.14		
NEO-FFI_Extraversion	-0.16*			0.04		
NEO-FFI_Offenheit für Neues	0.05			-0.12		
NEO-FFI_Verträglichkeit	-0.02			0.00		
NEO-FFI_Gewissenhaftigkeit	0.15			0.20*		
Geschlecht ¹	0.12	-0.04	-0.10	-0.03	-0.05	0.05
Alter ²	-0.07	0.02	0.12	-0.10	-0.06	-0.05
Schulabschluss	0.19*	0.13	0.04	-0.05	-0.02	0.06
Erfahrungen im Pflegebereich ²	0.00	0.08	0.13	-0.04	0.02	-0.12

Anmerkungen:

Korrelation nach Pearson. Einseitige Testung, * $p < 0.05$. ** $p < 0.01$.

d2-R= Aufmerksamkeits-Konzentrations-Test d2, revidierte Form (Brickenkamp, Schmidt-Atzert, & Liepmann, 2010). IST= Intelligenz-Struktur-Test 2000-R (Amthauer, Brocke, Liepmann, & Beauducel, 2001). Merkfähigkeit= WIT Skala Merkfähigkeit (Kersting, Althoff, & Jäger, 2008). Empathie= Situationsbezogener Empathiefragebogen (Holz-Ebeling & Steinmetz, 1994). EKF= Emotionale-Kompetenz-Fragebogen (Rindermann, 2007). EKF_EE= EKF-Skala Erkennen eigener Emotionen; EKF_EA= EKF-Skala Erkennen Emotionen anderer; EKF_RG= EKF-Skala Regulation und Kontrolle; EKF_EX= EKF-Skala Emotionale Expressivität. NEO-FFI= NEO-Fünf-Faktoren Inventar nach Costa und McCrae (Borkenau & Ostendorf, 1993).

¹ Kodierung: 0= männlich, 1= weiblich.

² in Jahren.

Anhang C. Prozentuale Verteilung der Probandenantworten auf die Antwortoptionen

Tabelle A-3: Angaben der Originalpersonen im Vergleich zur Häufigkeit der von den Probanden gewählten Antwortoptionen (ViTEIP-V2)

Item	AdO	Stichprobe 1				Stichprobe 2a				Stichprobe 2b			
		0	1	2	3	0	1	2	3	0	1	2	3
s1e1	3	11.6	32.1	42.9	13.4	27.7	35.0	28.5	8.8	48.4	21.3	23.8	6.6
s1e2	3	1.8	15.2	41.1	42.0	8.0	23.2	42.0	26.8	13.9	23.8	39.3	23.0
s1e4	0	49.6	27.4	15.9	7.1	23.4	29.2	29.2	18.3	50.8	20.5	14.8	13.9
s1e5	0	10.6	34.5	36.3	18.6	19.6	28.3	37.0	15.2	46.7	22.1	22.1	9.0
s1e6	2	0.9	8.0	25.7	65.5	0.7	5.1	16.8	77.4	2.5	3.3	14.8	79.5
s1e7	1	19.6	17.0	29.5	33.9	5.8	11.7	39.4	43.1	14.8	14.8	27.9	42.6
s1e8	2	54.5	31.3	12.5	1.8	42.7	25.7	19.9	11.8	66.4	14.8	10.7	8.2
s1e9	0	8.0	23.9	36.3	31.9	19.0	16.1	32.9	32.1	34.4	13.1	19.7	32.8
s2e2	0	57.1	33.0	8.0	1.8	52.9	24.6	18.1	4.4	70.5	12.3	11.5	5.7
s2e4	0	16.8	35.4	36.3	11.5	11.6	24.6	31.2	32.6	18.9	11.5	32.8	36.9
s2e5	0	18.6	50.4	25.7	5.3	42.7	36.0	17.7	3.7	64.8	17.2	8.2	9.8
s2e6	0	58.4	29.2	9.7	2.7	48.9	36.5	9.5	5.1	72.1	14.8	7.4	5.7
s2e8	1	2.7	28.3	44.3	24.8	5.1	28.3	40.6	26.1	27.9	13.1	27.9	31.2
s2e9	2	12.7	32.7	37.3	17.3	25.4	23.2	34.1	17.4	40.2	17.2	18.0	24.6
s2e10	2	65.2	17.9	13.4	3.6	62.8	16.8	11.7	8.8	75.4	7.4	6.6	10.7
s2e11	1	75.7	12.6	6.3	5.4	67.4	14.5	10.9	7.3	70.5	9.0	7.4	13.1
s2e12	3	8.9	28.3	33.6	29.2	5.8	21.7	37.0	35.5	23.0	15.6	27.1	34.4
s2e13	2	15.9	31.0	45.1	8.0	29.0	36.2	26.8	8.0	69.7	12.3	11.5	6.6
s2e14	1	48.2	29.5	19.6	2.7	50.7	21.7	20.3	7.3	67.2	9.0	14.8	9.0
s4e1	2	13.4	19.6	33.9	33.0	18.8	21.7	21.7	37.7	44.3	12.3	20.5	23.0
s4e2	3	0.9	4.4	23.0	71.7	2.2	7.3	15.2	75.4	11.5	6.6	15.6	66.4
s4e4	0	38.9	29.2	24.8	7.1	29.7	23.9	24.6	21.7	68.9	9.0	14.8	7.4
s4e6	3	0.9	5.3	13.3	80.5	0.0	2.9	9.5	87.6	4.1	0.8	18.9	76.2
s4e7	3	0.0	0.9	8.9	90.3	0.0	2.9	12.3	84.8	2.5	4.9	17.2	75.4
s4e8	1	7.1	32.7	42.5	17.7	8.0	20.4	38.7	32.9	18.9	27.9	32.0	21.3
s4e9	0	45.5	31.3	17.0	6.3	50.0	28.3	17.4	4.4	78.7	12.3	6.6	2.5
s4e14	0	35.4	27.4	17.7	19.5	27.0	16.8	20.4	35.8	59.0	15.6	13.9	11.5
s5e1	2	11.5	38.1	40.7	9.7	33.3	22.5	23.2	21.0	46.7	10.7	18.9	23.8
s5e2	2	0.9	13.3	54.0	31.9	2.9	12.4	43.1	41.6	9.0	9.8	29.5	51.6
s5e4	0	30.1	46.9	17.7	5.3	26.1	39.9	27.5	6.5	50.0	18.0	21.3	10.7
s5e5	0	21.4	41.1	29.5	8.0	23.0	43.7	28.9	4.4	45.1	18.9	20.5	15.6
s5e7	1	37.2	32.7	23.0	7.1	13.8	25.4	31.9	29.0	32.8	14.8	24.6	27.9
s5e14	3	19.5	22.1	41.6	16.8	20.9	27.6	25.4	26.1	44.3	17.2	16.4	22.1
s6e1	0	51.4	22.5	21.6	4.5	44.1	27.2	19.1	9.6	76.2	14.8	4.1	4.9
s6e4	0	56.3	35.7	8.0	0.0	45.6	28.7	17.7	8.1	73.8	10.7	13.1	2.5
s6e5	0	31.9	38.9	26.6	2.7	38.7	35.0	19.7	6.6	61.5	15.6	18.0	4.9
s6e6	3	6.2	24.8	37.2	31.9	1.5	13.8	34.8	50.0	9.8	13.9	29.5	46.7
s6e7	2	2.7	9.7	28.3	59.3	0.7	3.6	24.6	71.0	4.9	1.6	14.8	78.7
s6e14	3	12.4	14.2	31.9	41.6	5.3	9.1	22.7	62.9	22.1	13.9	24.6	39.3
s7e1	0	58.4	23.9	14.2	3.5	58.7	29.0	10.1	2.2	84.4	9.8	2.5	3.3
s7e2	0	22.3	42.0	30.4	5.4	26.8	33.3	26.1	13.8	49.2	23.0	18.0	9.8
s7e4	3	1.8	11.5	35.4	51.3	3.6	7.3	28.3	60.9	18.2	2.5	15.7	63.6
s7e5	3	0.9	0.9	17.7	80.5	1.5	2.2	18.8	77.5	8.2	0.8	7.4	83.6
s7e6	0	39.8	36.3	22.1	1.8	24.1	36.5	26.3	13.1	58.2	21.3	13.9	6.6
s7e7	0	45.5	28.6	22.3	3.6	28.5	33.6	24.8	13.1	60.7	13.9	15.6	9.8
s7e8	1	50.0	28.6	15.2	6.3	43.4	35.3	15.4	5.9	72.7	8.3	14.1	5.0
s7e9	0	12.4	29.2	33.6	24.8	18.1	23.2	29.0	29.7	31.2	18.0	25.4	25.4
s7e14	2	56.6	27.4	15.9	0.0	48.9	29.6	19.3	2.2	94.3	3.3	2.5	0.0

Tabelle wird fortgesetzt

Fortsetzung Tabelle A-3: Angaben der Originalpersonen im Vergleich zur Häufigkeit der von den Probanden gewählten Antwortoptionen (ViTEIP-V2)

Item	AdO	Stichprobe 1				Stichprobe 2a				Stichprobe 2b			
		0	1	2	3	0	1	2	3	0	1	2	3
s9e1	2	14.2	22.1	34.5	29.2	20.2	15.7	29.9	34.3	54.9	10.7	18.9	15.6
s9e2	2	8.0	18.6	31.9	41.6	2.9	11.7	16.8	68.6	13.9	15.6	22.1	48.4
s9e5	1	12.4	31.9	33.6	22.1	8.9	25.2	28.9	37.0	48.4	8.2	21.3	22.1
s9e6	0	39.6	34.2	18.9	7.2	13.9	21.2	35.8	29.2	50.8	13.9	23.8	11.5
s9e7	2	8.9	25.7	34.5	31.0	2.2	8.2	16.3	73.3	13.9	10.7	33.6	41.8
s9e8	2	12.5	25.9	33.9	27.7	10.9	21.7	37.0	30.4	49.2	11.5	16.4	23.0
s9e9	3	1.8	8.0	18.6	71.7	1.5	8.8	20.4	69.3	8.2	8.2	18.9	64.8
s9e11	0	60.2	12.4	20.4	7.1	52.9	10.3	16.2	20.6	73.0	4.1	7.4	15.6
S9e14	2	50.9	25.0	20.5	3.6	22.6	23.4	29.2	24.8	78.7	6.6	7.4	7.4
s9e15	2	8.3	14.8	25.0	51.9	10.2	19.0	23.4	47.5	13.1	17.2	14.8	54.9
s11e5	0	46.9	30.1	20.4	2.7	51.5	24.6	18.1	5.8	81.8	8.3	8.3	1.7
s11e10	3	0.9	1.8	15.0	82.3	0.7	0.7	10.9	87.7	2.5	0.0	6.6	91.0
s11e11	3	0.0	5.4	21.4	73.2	0.0	2.9	15.9	81.2	2.5	2.5	9.0	86.1
s11e12	2	43.8	28.6	23.2	4.5	29.4	33.8	25.7	11.0	67.2	13.1	17.2	2.5
s11e13	1	6.2	21.2	38.9	33.6	2.9	8.0	35.5	53.6	24.6	10.7	36.1	28.7
s11e14	1	38.1	22.1	31.0	8.9	15.3	16.1	37.2	31.4	43.4	11.5	31.2	13.9
s12e1	1	45.5	30.4	19.6	4.5	36.8	27.9	22.8	12.5	62.3	8.2	10.7	18.9
s12e2	1	30.1	28.3	29.2	12.4	21.0	32.6	29.0	17.4	48.4	11.5	14.8	25.4
s12e4	2	26.4	41.8	27.3	4.6	21.8	24.1	36.1	18.1	54.1	12.3	13.1	20.5
s12e5	0	21.2	26.6	41.6	10.6	14.1	26.7	42.2	17.0	41.0	10.7	23.0	25.4
s12e6	0	36.3	38.1	21.2	4.4	29.9	23.1	32.8	14.2	54.9	14.8	16.4	13.9
s12e7	0	30.4	32.1	25.9	11.6	17.4	29.0	33.3	20.3	47.5	9.8	20.5	22.1
s12e8	0	11.5	23.0	40.7	24.8	16.1	32.1	27.7	24.1	46.7	11.5	24.6	17.2
s12e9	1	17.0	36.6	26.8	19.6	33.1	24.3	20.6	22.1	43.4	11.5	20.5	24.6
s12e11	0	69.0	12.4	12.4	6.2	44.9	15.2	18.1	21.7	52.5	6.6	9.8	31.2
s12e14	1	51.3	25.7	16.8	6.2	48.2	22.6	16.1	13.1	82.0	6.6	4.9	6.6
s13e1	1	54.0	31.0	13.3	1.8	63.7	22.2	6.7	7.4	80.3	10.7	4.9	4.1
s13e2	2	14.4	46.9	30.6	8.1	11.7	27.0	38.7	22.6	33.6	25.4	29.5	11.5
s13e5	0	31.3	43.8	21.4	3.6	29.2	30.7	30.7	9.5	65.6	11.5	18.0	4.9
s13e6	3	5.4	19.6	33.9	41.1	0.0	5.1	23.9	71.0	6.6	9.0	18.9	65.6
s13e7	3	1.8	12.4	31.0	54.9	0.7	5.8	18.8	74.6	2.5	4.1	16.4	77.1
s13e10	0	46.0	26.6	20.4	7.1	44.9	28.3	15.2	11.6	59.0	12.3	17.2	11.5
s13e11	1	25.7	18.6	35.4	20.4	26.3	13.9	26.3	33.6	50.0	7.4	16.4	26.2
s13e12	0	60.6	26.6	10.1	2.8	51.8	28.5	15.3	4.4	82.8	7.4	7.4	2.5
s13e14	0	63.7	19.5	9.7	7.1	41.3	25.4	21.7	11.6	91.0	4.1	3.3	1.6
s15e1	0	47.8	33.6	13.3	5.3	45.6	27.9	16.9	9.6	72.7	16.5	5.8	5.0
s15e2	0	62.0	24.8	9.7	3.5	42.3	27.0	20.4	10.2	74.4	5.0	15.7	5.0
s15e4	2	4.4	9.7	42.5	43.4	2.2	9.6	33.1	55.2	14.9	12.4	19.0	53.7
s15e5	2	8.1	10.8	41.4	39.6	8.1	18.4	39.7	33.8	38.0	11.6	27.3	23.1
s15e6	2	1.8	7.1	23.9	67.3	1.5	5.8	12.4	80.3	9.1	8.3	14.1	68.6
s15e7	0	27.7	36.6	24.1	11.6	16.1	18.3	28.5	37.2	46.3	8.3	24.0	21.5
s15e9	0	18.8	30.4	35.7	15.2	19.0	25.6	30.7	24.8	41.3	18.2	22.3	18.2
s16e1	3	6.2	18.6	46.0	29.2	22.1	34.6	27.2	16.2	49.2	9.0	22.1	19.7
s16e2	3	0.0	8.9	28.6	62.5	1.5	7.3	31.9	59.4	8.2	9.0	14.8	68.0
s16e5	0	31.9	31.9	29.2	7.1	21.2	31.4	35.8	11.7	67.2	9.8	13.9	9.0
s16e6	0	59.8	28.6	8.9	2.7	24.1	35.0	24.8	16.1	56.6	13.1	22.1	8.2
s16e7	1	15.9	20.4	43.4	20.4	4.4	10.2	35.0	50.4	20.5	12.3	21.3	45.9
s16e8	0	52.7	31.3	13.4	2.7	39.7	36.0	16.2	8.1	70.5	13.9	8.2	7.4
s16e9	0	25.9	33.9	30.4	9.8	20.4	24.8	36.5	18.3	44.3	19.7	13.1	23.0
s16e14	1	22.3	25.0	38.4	14.3	23.4	19.7	22.6	34.3	63.1	6.6	14.8	15.6

Anmerkungen:

AdO= Angabe der Originalperson. 0= relative Häufigkeit der gewählten Antwortoption „Emotion nicht vorhanden“; 1= relative Häufigkeit der gewählten Antwortoption „Emotion schwach vorhanden“; 2= relative Häufigkeit der gewählten Antwortoption „Emotion mittel vorhanden“; 3= relative Häufigkeit der gewählten Antwortoption „Emotion stark vorhanden“. Angaben in Prozent; entspricht der Gewichtung bei Konsens-Anteil.

Tabelle A-4: Angaben der Originalpersonen im Vergleich zur Häufigkeit der von den Probanden gewählten Antwortoptionen (TEMINT)

Item	AdO	Stichprobe 1			Stichprobe 2a			Stichprobe 2b		
		0	1	2	0	1	2	0	1	2
s1e3	2	0.9	8.8	90.3	0.0	8.0	92.0	1.7	12.4	86.0
s1e4	2	2.7	14.2	83.2	2.9	33.3	63.8	6.6	30.6	62.8
s1e5	2	19.5	55.8	24.8	9.4	41.3	49.3	15.1	44.5	40.3
s1e6	2	0.0	4.4	95.6	0.0	6.5	93.5	4.1	11.6	84.3
s1e7	0	100.0	0.0	0.0	97.1	1.5	1.5	100.0	0.0	0.0
s1e8	0	99.1	0.9	0.0	97.8	2.2	0.0	100.0	0.0	0.0
s1e9	2	26.5	36.3	37.2	25.4	27.5	47.1	30.5	37.3	32.2
s1e10	2	36.3	54.9	8.8	40.9	45.3	13.9	69.7	26.1	4.2
s2e1	0	93.8	6.2	0.0	81.2	14.5	4.3	83.2	12.6	4.2
s2e2	0	92.0	7.1	0.9	75.4	17.4	7.2	82.2	12.7	5.1
s2e3	2	14.2	51.3	34.5	9.5	50.4	40.1	14.9	48.8	36.4
s2e4	2	1.8	41.6	56.6	6.6	46.0	47.4	14.0	37.2	48.8
s2e7	2	4.5	18.8	76.8	11.6	13.0	75.4	11.7	13.3	75.0
s2e8	2	16.8	30.1	53.1	13.8	27.5	58.7	12.6	28.6	58.8
s2e9	2	14.2	22.1	63.7	17.4	30.4	52.2	19.3	30.3	50.4
s3e1	2	52.2	38.1	9.7	50.0	41.2	8.8	55.0	37.5	7.5
s3e2	2	1.8	23.9	74.3	1.4	18.1	80.4	4.9	16.4	78.7
s3e3	2	36.6	44.6	18.8	15.2	49.3	35.5	15.7	57.0	27.3
s3e4	2	3.5	27.4	69.0	2.9	31.2	65.9	4.1	30.3	65.6
s3e5	1	58.0	37.5	4.5	38.7	46.0	15.3	48.8	43.0	8.3
s3e6	2	19.5	53.1	27.4	19.9	44.9	35.3	29.2	45.8	25.0
s3e7	0	100.0	0.0	0.0	98.5	0.7	0.7	100.0	0.0	0.0
s3e8	0	96.5	2.7	0.9	92.8	5.1	2.2	95.9	4.1	0.0
s3e9	0	97.3	1.8	0.9	87.0	10.9	2.2	95.8	4.2	0.0
s4e1	2	2.7	58.4	38.9	19.6	54.3	26.1	21.3	47.5	31.1
s4e2	2	2.7	15.9	81.4	0.7	15.2	84.1	1.6	13.1	85.2
s4e3	1	61.9	32.7	5.3	34.3	55.5	10.2	38.5	54.9	6.6
s4e4	2	9.7	63.7	26.5	8.7	52.2	39.1	12.4	47.9	39.7
s4e5	1	29.5	61.6	8.9	16.7	53.6	29.7	24.8	45.5	29.8
s4e7	0	99.1	0.9	0.0	98.6	1.4	0.0	99.2	0.8	0.0
s4e8	2	86.7	12.4	0.9	86.2	11.6	2.2	88.5	10.7	0.8
s4e10	2	45.1	52.2	2.7	43.5	48.6	8.0	64.8	32.0	3.3
s5e1	0	99.1	0.9	0.0	98.6	1.4	0.0	96.7	3.3	0.0
s5e2	0	99.1	0.9	0.0	97.1	2.9	0.0	99.2	0.8	0.0
s5e5	0	93.8	6.2	0.0	95.7	2.9	1.4	96.7	3.3	0.0
s5e6	0	86.7	12.4	0.9	87.7	11.6	0.7	95.1	4.9	0.0
s5e7	2	0.9	2.7	96.5	2.2	4.4	93.4	0.8	4.9	94.3
s5e10	2	0.0	7.1	92.9	0.0	5.8	94.2	0.8	9.0	90.2
s6e1	2	0.0	16.8	83.2	4.4	29.2	66.4	2.5	27.0	70.5
s6e2	2	1.8	39.8	58.4	3.6	47.4	48.9	4.1	38.5	57.4
s6e3	1	77.9	21.2	0.9	67.6	25.7	6.6	62.5	35.8	1.7
s6e4	2	18.9	68.5	12.6	13.1	62.8	24.1	6.6	61.5	32.0
s6e7	0	100.0	0.0	0.0	99.3	0.7	0.0	98.4	0.0	1.6
s6e9	0	100.0	0.0	0.0	92.7	5.8	1.5	92.6	7.4	0.0
s7e1	2	2.7	34.5	62.8	8.0	28.3	63.8	9.8	32.0	58.2
s7e2	2	3.5	18.6	77.9	2.2	13.0	84.8	0.8	20.5	78.7
s7e4	2	42.5	53.1	4.4	32.6	50.0	17.4	29.5	55.7	14.8
s7e6	0	89.4	9.7	0.9	89.9	8.7	1.4	91.0	8.2	0.8
s7e7	0	98.2	0.9	0.9	97.1	1.4	1.4	99.2	0.8	0.0
s7e9	0	92.9	7.1	0.0	90.6	8.0	1.4	95.1	4.9	0.0
s8e3	2	1.8	16.8	81.4	1.4	13.8	84.8	5.0	9.2	85.7
s8e4	2	0.0	21.2	78.8	2.9	21.0	76.1	1.7	29.7	68.6
s8e5	2	0.0	15.9	84.1	0.0	7.2	92.8	1.7	12.7	85.6
s8e9	2	15.0	33.6	51.3	15.9	34.1	50.0	28.6	32.8	38.7
s8e10	2	1.8	25.7	72.6	6.6	14.6	78.8	16.8	26.9	56.3

Tabelle wird fortgesetzt

Fortsetzung Tabelle A-4: Angaben der Originalpersonen im Vergleich zur Häufigkeit der von den Probanden gewählten Antwortoptionen (TEMINT)

Item	AdO	Stichprobe 1			Stichprobe 2a			Stichprobe 2b		
		0	1	2	0	1	2	0	1	2
s9e2	2	3.5	58.4	38.1	0.7	34.1	65.2	4.1	36.9	59.0
s9e3	2	11.5	45.1	43.4	5.8	33.3	60.9	3.3	37.2	59.5
s9e4	2	3.5	45.1	51.3	1.4	39.1	59.4	1.6	42.6	55.7
s9e5	2	1.8	39.8	58.4	0.0	17.5	82.5	6.6	23.8	69.7
s9e7	0	96.5	3.5	0.0	97.1	2.9	0.0	95.1	4.1	0.8
s9e9	2	31.0	61.1	8.0	32.6	46.4	21.0	45.9	44.3	9.8
s10e2	0	100.0	0.0	0.0	98.6	0.7	0.7	99.2	0.8	0.0
s10e3	0	100.0	0.0	0.0	100.0	0.0	0.0	99.2	0.8	0.0
s10e5	0	100.0	0.0	0.0	100.0	0.0	0.0	98.3	1.7	0.0
s10e7	2	0.0	0.0	100.0	0.7	0.0	99.3	1.6	0.0	98.4
s10e8	2	0.0	3.5	96.5	0.7	4.3	94.9	1.6	2.5	95.9
s10e9	2	12.4	42.5	45.1	13.8	41.3	44.9	18.0	41.0	41.0
s10e10	1	14.2	68.1	17.7	5.1	39.4	55.5	13.9	51.6	34.4
s11e1	1	54.0	40.7	5.3	54.0	40.1	5.8	52.9	40.5	6.6
s11e2	2	5.3	62.8	31.9	11.6	33.3	55.1	6.6	44.3	49.2
s11e3	2	0.9	26.5	72.6	3.6	17.4	79.0	7.4	18.9	73.8
s11e4	2	0.9	8.8	90.3	3.6	15.2	81.2	5.7	14.8	79.5
s11e7	0	97.3	2.7	0.0	91.3	5.1	3.6	95.9	2.5	1.6
s11e8	0	96.5	3.5	0.0	95.7	3.6	0.7	92.6	7.4	0.0
s11e9	1	31.0	54.0	15.0	42.8	32.6	24.6	45.1	37.7	17.2
s12e1	0	100.0	0.0	0.0	98.6	0.7	0.7	98.3	1.7	0.0
s12e2	0	100.0	0.0	0.0	100.0	0.0	0.0	99.2	0.8	0.0
s12e3	1	80.5	19.5	0.0	81.2	15.9	2.9	76.2	23.0	0.8
s12e4	2	62.8	35.4	1.8	65.2	31.2	3.6	68.0	30.3	1.6
s12e5	0	100.0	0.0	0.0	99.3	0.7	0.0	99.2	0.8	0.0
s12e6	0	99.1	0.9	0.0	99.3	0.7	0.0	99.2	0.0	0.8
s12e7	2	0.0	3.5	96.5	0.0	0.0	100.0	0.0	0.8	99.2
s12e8	2	1.8	13.3	85.0	0.0	8.7	91.3	1.6	9.8	88.5
s12e9	2	0.9	9.7	89.4	3.6	21.7	74.6	6.6	10.7	82.8
s12e10	1	81.4	18.6	0.0	48.6	34.1	17.4	47.9	30.6	21.5

Anmerkungen:

AdO= Angabe der Originalperson. 0= gewählte Antwortoption „Emotion nicht oder schwach vorhanden“; 1= gewählte Antwortoption „Emotion schwach bis mittel vorhanden“; 2= gewählte Antwortoption „Emotion stark bis sehr stark vorhanden“; Angaben in Prozent; entspricht der Gewichtung bei Konsens-Anteil.

Anhang D. Deskriptive Statistiken*Tabelle A-5: Deskriptive Statistiken der eingesetzten Testverfahren*

Testverfahren	Stichprobe	N	M	SD	Mittelwertvergleiche der Stichproben:		
					1 und 2a	1 und 2b	2a und 2b
Test d2-R (Konzentrationsleistung)	1	113	178.43	38.25	$t=8.50^*$	$t=8.51^*$	$t=-0.34$
	2a	136	137.15	38.06	$(p=0.00)$	$(p=0.00)$	$(p=0.73)$
	2b	120	138.69	32.92	$d=0.71$	$d=0.76$	$d=0.03$
EKF-Selbst Gesamt	1	111	0.42	0.66	$t=1.80$	$t=2.17$	$t=0.36$
	2a	136	0.26	0.69	$(p=0.07)$	$(p=0.03)$	$(p=0.72)$
	2b	116	0.23	0.63	$d=0.01$	$d=0.02$	$d=0.00$
Erkennen eigener Emotionen	1	111	3.58	0.64	$t=2.30^*$	$t=2.30^*$	$t=-0.01$
	2a	136	3.39	0.66	$(p=0.02)$	$(p=0.02)$	$(p=0.99)$
	2b	118	3.39	0.62	$d=0.02$	$d=0.02$	$d=0.00$
Erkennen Emotionen anderer	1	112	3.95	0.50	$t=0.94$	$t=1.89$	$t=0.88$
	2a	136	3.88	0.54	$(p=0.35)$	$(p=0.06)$	$(p=0.38)$
	2b	119	3.83	0.44	$d=0.01$	$d=0.01$	$d=0.00$
Regulation und Kontrolle	1	112	3.41	0.69	$t=-1.28$	$t=-1.93$	$t=-0.63$
	2a	136	3.52	0.65	$(p=0.20)$	$(p=0.05)^1$	$(p=0.53)$
	2b	119	3.57	0.52	$d=0.01$	$d=0.01$	$d=0.00$
Emotionale Expressivität	1	111	3.43	0.76	$t=2.69^*$	$t=3.68^*$	$t=0.42$
	2a	136	3.18	0.72	$(p=0.01)$	$(p=0.00)$	$(p=0.31)$
	2b	117	3.09	0.64	$d=0.02$	$d=0.03$	$d=0.01$
NEO-FFI Neurotizismus	1	113	1.93	0.73			
NEO-FFI Extraversion	1	113	2.38	0.59			
NEO-FFI Offenheit für Neues	1	113	2.78	0.57			
NEO-FFI Verträglichkeit	1	113	2.72	0.50			
NEO-FFI Gewissenhaftigkeit	1	113	2.47	0.63			
IST Verbales Schlussfolgern	1	113	39.91	6.14			
IST Figurales Schlussfolgern	1	113	32.32	7.99			
IST Verbale Merkfähigkeit	1	113	8.01	2.27			
WIT-2	2a	135	12.33	3.60			$t=-0.92$
	2b	121	12.75	3.70			$(p=0.36)$ $d=0.03$
EKF-Fremd Schüler	2a	30 ²	0.25	0.71			$t=1.02$
	2b	26 ²	0.06	0.73			$(p=0.31)$ $d=0.04$
Situationsbezogener Empathiefragebogen	2a	135	84.87	11.09			$t=0.42$
	2b	107	84.27	10.91			$(p=0.67)$ $d=0.03$

Anmerkungen:

N= Anzahl der Probanden; M= Mittelwert; SD= Standardabweichung. Mittelwertvergleiche= t-Test für unabhängige Stichproben, zweiseitige Testung zwischen je zwei Stichproben. d = Effektstärke.

VITEIP= Videobasierter Test zur Erfassung der Emotionalen Intelligenz im Pflegeberuf. TEMINT= Test zur Erfassung der Emotionalen Intelligenz (Schmidt-Atzert & Bühner, 2002). d2-R= Aufmerksamkeits-Konzentrations-Test d2, revidierte Form (Brickenkamp, Schmidt-Atzert, & Liepmann, 2010). NEO-FFI= NEO-Fünf-Faktoren Inventar nach Costa und McCrae (Borkenau & Ostendorf, 1993). IST-2000-R= Intelligenz-Struktur-Test 2000-R (Amthauer, Brocke, Liepmann, & Beauducel, 2001). WIT-2= Der Wilde-Intelligenztest (Kersting, Althoff, & Jäger, 2008). EKF= Emotionale-Kompetenz-Fragebogen (Rindermann, 2007). Situationsbezogener Empathiefragebogen= situationsbezogenen Fragebogen zur Erfassung von Empathie (Holz-Ebeling & Steinmetz, 1994).

Einige Zellen bleiben leer, da nicht alle Verfahren in allen Untersuchungen zum Einsatz kamen.

¹ Levene-Test der Varianzgleichheit auf dem 0.05 Niveau signifikant.

² Kleines N aufgrund geringer Rücklaufquote, Daten sind damit nur bedingt aussagekräftig.

*Auf dem Niveau 0.05 signifikant.

Anhang E. Faktorenanalysen ViTEIP-V2 und TEMINT

ViTEIP-V2, Stichprobe 1, Scoring Target-Übereinstimmung

Dateneignung: KMO=0.66; Bartlett-Test signifikant ($p=0.00$)

Tabelle A-6: Strukturmatrix einer Hauptkomponentenanalyse der Situationen des ViTEIP-V2, Stichprobe 1, Target-Übereinstimmung

Sit.	Person	Ereignis	Dominante Emotionen laut Angabe der Originalperson ¹	Komponente ²				
				1	2	3	4	5
1	Frau Klein (A)	Zu Bett gebracht werden	Abneigung, Ärger , Trauer, Schuld	0.21	0.69	0.28	0.52	0.31
2	Frau Schmidt (A)	Ein Glas zerbricht	Dankbarkeit , Scham, Freude, Zuneigung	0.05	0.05	0.71	-0.17	0.02
4	Frau Fischer (A)	Abgebrochener Besuch	Ärger, Trauer, Enttäuschung , Abneigung	0.09	0.23	0.60	0.40	0.06
5	Frau Schubert (R)	Unerwartete Nachvisite	Überraschung , Abneigung, Ärger	0.16	0.73	-0.09	-0.08	-0.31
6	Frau Albrecht (R)	Milchunverträglichkeit	Trauer, Überraschung , Enttäuschung	0.47	0.20	0.38	0.12	-0.40
7	Frau Hoffmann (A)	Bestellung	Angst, Unruhe , Überraschung	0.69	0.13	0.12	-0.06	-0.11
9	Frau Schubert (R)	Plötzlicher Durchfall	Scham , Abneigung, Enttäuschung, Ärger, Ekel, Schuld, Überraschung	0.20	0.65	0.38	0.12	0.11
11	Frau Fischer (A)	Wellensittich gefunden	Freude, Stolz , Dankbarkeit	0.10	0.01	0.09	0.02	0.81
12	Frau Krüger (R)	Sturz ohne Korsett	Angst	0.65	0.49	0.06	0.35	0.04
13	Frau Schubert (R)	Ausbleibender Anruf	Trauer, Enttäuschung , Ärger	0.03	0.06	0.01	0.76	-0.02
15	Frau Ludwig (R)	Plötzliches Weinen	Angst, Unruhe, Trauer	0.67	0.04	0.01	0.14	0.39
16	Frau Fischer (A)	Bitte um Toilettenpapier	Abneigung, Ärger	0.53	0.29	0.50	0.50	0.12
Erklärte Varianz in % (pro Komponente)				20.66	10.82	9.90	9.30	8.61

Anmerkungen:

Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Promax mit Kaiser-Normalisierung.

(A)= Die Person ist Bewohnerin der Altenhilfeeinrichtung; (R)= Die Person ist Patientin in der Rehaklinik.

¹ **Fettgedruckt**: Emotionen, die als „stark vorhanden“ eingestuft wurden; Normal gedruckt: Emotionen, die als „mittel vorhanden“ eingestuft wurden.

² Aus Gründen der Übersichtlichkeit sind Ladungen >0.30 **fettgedruckt**.

Tabelle A-7: Komponentenkorrelationsmatrix

Komponente	1	2	3	4	5
1					
2	0.30				
3	0.24	0.22			
4	0.24	0.29	0.23		
5	0.05	0.01	0.07	0.21	

Anmerkungen:

Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Promax mit Kaiser-Normalisierung.

ViTEIP-V2, Stichprobe 1, Scoring Target-TrefferDateneignung: KMO=0.69; Bartlett-Test signifikant ($p=0.00$)*Tabelle A-8: Strukturmatrix einer Hauptkomponentenanalyse der Situationen des ViTEIP-V2, Stichprobe 1, Target-Treffer*

Sit.	Person	Ereignis	Dominante Emotionen laut Angabe der Originalperson ¹	Komponente ²			
				1	2	3	4
1	Frau Klein (A)	Zu Bett gebracht werden	Abneigung, Ärger , Trauer, Schuld	0.72	0.12	0.24	-0.09
2	Frau Schmidt (A)	Ein Glas zerbricht	Dankbarkeit , Scham, Freude, Zuneigung	0.24	-0.01	0.08	0.63
4	Frau Fischer (A)	Abgebrochener Besuch	Ärger, Trauer, Enttäuschung , Abneigung	0.65	0.24	-0.04	0.12
5	Frau Schubert (R)	Unerwartete Nachvisite	Überraschung , Abneigung, Ärger	0.36	0.19	0.21	-0.67
6	Frau Albrecht (R)	Milchunverträglichkeit	Trauer, Überraschung , Enttäuschung	0.44	0.64	0.05	-0.06
7	Frau Hoffmann (A)	Bestellung	Angst, Unruhe , Überraschung	0.17	0.82	0.12	-0.07
9	Frau Schubert (R)	Plötzlicher Durchfall	Scham , Abneigung, Enttäuschung, Ärger, Ekel, Schuld, Überraschung	0.61	0.30	0.21	-0.42
11	Frau Fischer (A)	Wellensittich gefunden	Freude, Stolz , Dankbarkeit	0.27	0.10	-0.51	0.33
12	Frau Krüger (R)	Sturz ohne Korsett	Angst	0.25	0.03	0.80	0.03
13	Frau Schubert (R)	Ausbleibender Anruf	Trauer, Enttäuschung , Ärger	0.05	0.36	-0.05	0.43
15	Frau Ludwig (R)	Plötzliches Weinen	Angst, Unruhe, Trauer	0.27	0.43	0.57	-0.04
16	Frau Fischer (A)	Bitte um Toilettenpapier	Abneigung, Ärger	0.63	0.47	0.34	0.16
Erklärte Varianz in % (pro Komponente)				22.13	11.65	9.65	8.80

Anmerkungen:

Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Promax mit Kaiser-Normalisierung.

(A)= Die Person ist Bewohnerin der Altenhilfeeinrichtung; (R)= Die Person ist Patientin in der Rehaklinik.

¹ **Fettgedruckt:** Emotionen, die als „stark vorhanden“ eingestuft wurden; Normal gedruckt: Emotionen, die als „mittel vorhanden“ eingestuft wurden.² Aus Gründen der Übersichtlichkeit sind Ladungen >0.30 **fettgedruckt**.*Tabelle A-9: Komponentenkorrelationsmatrix*

Komponente	1	2	3	4
1				
2	0.38			
3	0.22	0.15		
4	-0.01	-0.01	-0.14	

Anmerkungen:

Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Promax mit Kaiser-Normalisierung.

ViTEIP-V2, Stichprobe 1, Scoring Konsens-AnteilDateneignung: KMO=0.61; Bartlett-Test signifikant ($p=0.00$)*Tabelle A-10: Strukturmatrix einer Hauptkomponentenanalyse der Situationen des ViTEIP-V2, Stichprobe 1, Konsens-Anteil*

Sit.	Person	Ereignis	Dominante Emotionen laut Angabe der Originalperson ¹	Komponente ²				
				1	2	3	4	5
1	Frau Klein (A)	Zu Bett gebracht werden	Abneigung, Ärger , Trauer, Schuld	-0.12	0.02	0.71	-0.07	-0.22
2	Frau Schmidt (A)	Ein Glas zerbricht	Dankbarkeit , Scham, Freude, Zuneigung	0.21	0.08	-0.04	0.73	0.20
4	Frau Fischer (A)	Abgebrochener Besuch	Ärger, Trauer, Enttäuschung , Abneigung	0.05	0.04	-0.02	0.09	0.84
5	Frau Schubert (R)	Unerwartete Nachvisite	Überraschung , Abneigung, Ärger	0.67	0.09	0.13	0.26	-0.25
6	Frau Albrecht (R)	Milchunverträglichkeit	Trauer, Überraschung , Enttäuschung	0.22	0.71	0.18	-0.18	0.23
7	Frau Hoffmann (A)	Bestellung	Angst, Unruhe , Überraschung	0.32	0.53	0.22	0.25	-0.08
9	Frau Schubert (R)	Plötzlicher Durchfall	Scham , Abneigung, Enttäuschung, Ärger, Ekel, Schuld, Überraschung	0.26	0.07	0.65	0.11	0.23
11	Frau Fischer (A)	Wellensittich gefunden	Freude, Stolz , Dankbarkeit	0.03	0.31	0.42	0.59	-0.17
12	Frau Krüger (R)	Sturz ohne Korsett	Angst	0.77	0.19	0.03	0.01	0.15
13	Frau Schubert (R)	Ausbleibender Anruf	Trauer, Enttäuschung , Ärger	0.48	0.15	0.14	-0.12	0.35
15	Frau Ludwig (R)	Plötzliches Weinen	Angst, Unruhe, Trauer	0.03	0.75	-0.07	0.21	-0.07
16	Frau Fischer (A)	Bitte um Toilettenpapier	Abneigung, Ärger	0.47	0.21	0.53	-0.50	0.19
Erklärte Varianz in % (pro Komponente)				17.69	11.68	10.63	9.63	8.94

Anmerkungen:

Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Promax mit Kaiser-Normalisierung.

(A)= Die Person ist Bewohnerin der Altenhilfeeinrichtung; (R)= Die Person ist Patientin in der Rehaklinik.

¹ **Fettgedruckt:** Emotionen, die als „stark vorhanden“ eingestuft wurden; Normal gedruckt: Emotionen, die als „mittel vorhanden“ eingestuft wurden.² Aus Gründen der Übersichtlichkeit sind Ladungen >0.30 **fettgedruckt**.*Tabelle A-11: Komponentenkorrelationsmatrix*

Komponente	1	2	3	4	5
1					
2	0.23				
3	0.24	0.20			
4	-0.02	0.08	-0.06		
5	0.18	0.05	0.01	-0.18	

Anmerkungen:

Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Promax mit Kaiser-Normalisierung.

ViTEIP-V2, Stichprobe 2a, Scoring Target-ÜbereinstimmungDateneignung: KMO=0.68; Bartlett-Test signifikant ($p=0.00$)*Tabelle A-12: Strukturmatrix einer Hauptkomponentenanalyse der Situationen des ViTEIP-V2, Stichprobe 2a, Target-Übereinstimmung*

Sit.	Person	Ereignis	Dominante Emotionen laut Angabe der Originalperson ¹	Komponente ²			
				1	2	3	4
1	Frau Klein (A)	Zu Bett gebracht werden	Abneigung, Ärger , Trauer, Schuld	-0.07	0.56	0.09	0.02
2	Frau Schmidt (A)	Ein Glas zerbricht	Dankbarkeit , Scham, Freude, Zuneigung	0.30	-0.14	0.69	0.14
4	Frau Fischer (A)	Abgebrochener Besuch	Ärger, Trauer, Enttäuschung , Abneigung	0.55	0.17	0.04	-0.17
5	Frau Schubert (R)	Unerwartete Nachvisite	Überraschung , Abneigung, Ärger	0.03	0.31	0.78	-0.02
6	Frau Albrecht (R)	Milchunverträglichkeit	Trauer, Überraschung , Enttäuschung	0.58	-0.07	0.04	-0.09
7	Frau Hoffmann (A)	Bestellung	Angst, Unruhe , Überraschung	0.43	0.23	0.28	0.70
9	Frau Schubert (R)	Plötzlicher Durchfall	Scham , Abneigung, Enttäuschung, Ärger, Ekel, Schuld, Überraschung	0.30	0.75	0.25	0.02
11	Frau Fischer (A)	Wellensittich gefunden	Freude, Stolz , Dankbarkeit	0.33	0.13	0.14	-0.77
12	Frau Krüger (R)	Sturz ohne Korsett	Angst	0.62	0.34	0.34	0.23
13	Frau Schubert (R)	Ausbleibender Anruf	Trauer, Enttäuschung , Ärger	0.18	0.60	-0.09	0.11
15	Frau Ludwig (R)	Plötzliches Weinen	Angst, Unruhe, Trauer	0.74	0.25	0.25	0.21
16	Frau Fischer (A)	Bitte um Toilettenpapier	Abneigung, Ärger	0.51	0.59	0.33	-0.14
Erklärte Varianz in % (pro Komponente)				22.05	11.58	10.63	9.23

Anmerkungen:

Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Promax mit Kaiser-Normalisierung.

(A)= Die Person ist Bewohnerin der Altenhilfeeinrichtung; (R)= Die Person ist Patientin in der Rehaklinik.

¹ **Fettgedruckt:** Emotionen, die als „stark vorhanden“ eingestuft wurden; Normal gedruckt: Emotionen, die als „mittel vorhanden“ eingestuft wurden.² Aus Gründen der Übersichtlichkeit sind Ladungen >0.30 **fettgedruckt**.*Tabelle A-13: Komponentenkorrelationsmatrix*

Komponente	1	2	3	4
1				
2	0.29			
3	0.32	0.23		
5	0.07	0.03	0.09	

Anmerkungen:

Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Promax mit Kaiser-Normalisierung.

TEMINT, Stichprobe 1, Scoring Target-ÜbereinstimmungDateneignung: KMO=0.72; Bartlett-Test signifikant ($p=0.00$)*Tabelle A-14: Strukturmatrix einer Hauptkomponentenanalyse der Situationen des TEMINT, Stichprobe 1, Target-Übereinstimmung*

Sit.	Person	Ereignis	Dominante Emotion ¹	Komponente ²			
				1	2	3	4
1	Informatikerin, 30 Jahre	Katze vergiftet	Unruhe, Zuneigung, Trauer, Angst, Überraschung, Schuld	0.64	0.42	0.19	0.18
2	Zahnarzthelferin, 23 Jahre	Geburt	Angst, Unruhe, Stolz, Freude, Zuneigung	0.34	0.59	-0.09	0.60
3	Student, 30 Jahre	Computerproblem	Abneigung, Ärger, Angst, Unruhe, Schuld, Trauer	0.37	0.57	0.24	-0.33
4	Beamten, 22 Jahre	Streit mit Kollegin	Abneigung, Unruhe, Ärger, Stolz, Überraschung, Trauer, Angst	0.63	0.40	0.46	-0.21
5	Sachbearbeiter, 29 Jahre	Zufälliges Treffen	Freude, Überraschung	-0.07	-0.13	0.14	0.76
6	Lagerist, 34 Jahre	Unfreundlicher Kunde	Abneigung, Ärger, Unruhe, Angst	0.61	0.04	0.67	-0.11
7	Diplom-Ingenieur, 69 Jahre	Zeitungsmeldung	Abneigung, Ärger, Unruhe	-0.07	0.04	0.72	0.07
8	Lehrerin, 58 Jahre	Erkrankte Schwägerin	Angst, Unruhe, Trauer, Zuneigung, Überraschung	0.31	0.75	0.17	0.18
9	Beamter, 44 Jahre	Eheprobleme	Ärger, Angst, Trauer, Unruhe, Zuneigung	0.31	0.76	0.10	-0.11
10	Schüler, 17 Jahre	Gewonnenes Fußballspiel	Freude, Zuneigung, Stolz, Überraschung	0.64	0.22	0.02	-0.13
11	Angestellte, 31 Jahre	Ausbleibender Anruf	Ärger, Angst, Unruhe, Abneigung, Zuneigung	0.29	0.35	0.55	0.16
12	Polizist, 24 Jahre	Frisch verheiratet	Unruhe, Freude, Stolz, Zuneigung, Angst, Überraschung	0.66	0.29	0.04	0.10
Erklärte Varianz in % (pro Komponente)				26.10	11.36	9.69	8.91

Anmerkungen:

Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Promax mit Kaiser-Normalisierung.

(A)= Die Person ist Bewohnerin der Altenhilfeeinrichtung; (R)= Die Person ist Patientin in der Rehaklinik.

¹ **Fettgedruckt:** Emotionen, die als „stark vorhanden“ eingestuft wurden; Normal gedruckt: Emotionen, die als „mittel vorhanden“ eingestuft wurden.² Aus Gründen der Übersichtlichkeit sind Ladungen >0.30 **fettgedruckt**.*Tabelle A-15: Komponentenkorrelationsmatrix*

Komponente	1	2	3	4
1				
2	0.43			
3	0.31	0.10		
4	-0.04	0.11	-0.10	

Anmerkungen:

Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Promax mit Kaiser-Normalisierung.

TEMINT, Stichprobe 1, Scoring Target-TrefferDateneignung: KMO=0.73; Bartlett-Test signifikant ($p=0.00$)*Tabelle A-16: Strukturmatrix einer Hauptkomponentenanalyse der Situationen des TEMINT, Stichprobe 1, Target-Treffer*

Sit.	Person	Ereignis	Dominante Emotion ¹	Komponente ²				
				1	2	3	4	5
1	Informatikerin, 30 Jahre	Katze vergiftet	Unruhe, Zuneigung, Trauer, Angst, Überraschung, Schuld	0.41	0.64	0.43	-0.12	0.14
2	Zahnarzthelferin, 23 Jahre	Geburt	Angst, Unruhe, Stolz, Freude, Zuneigung	0.66	0.22	0.28	-0.17	0.40
3	Student, 30 Jahre	Computerproblem	Abneigung, Ärger, Angst, Unruhe, Schuld, Trauer	0.52	0.11	0.35	0.39	-0.51
4	Beamten, 22 Jahre	Streit mit Kollegin	Abneigung, Unruhe, Ärger, Stolz, Überraschung, Trauer, Angst	0.43	0.77	0.23	0.12	-0.05
5	Sachbearbeiter, 29 Jahre	Zufälliges Treffen	Freude, Überraschung	-0.06	0.03	-0.02	0.12	0.80
6	Lagerist, 34 Jahre	Unfreundlicher Kunde	Abneigung, Ärger, Unruhe, Angst	0.05	0.76	0.12	0.29	0.02
7	Diplom-Ingenieur, 69 Jahre	Zeitungsmeldung	Abneigung, Ärger, Unruhe	-0.04	0.15	-0.17	0.71	-0.03
8	Lehrerin, 58 Jahre	Erkrankte Schwägerin	Angst, Unruhe, Trauer, Zuneigung, Überraschung	0.73	0.21	0.33	0.23	-0.05
9	Beamter, 44 Jahre	Eheprobleme	Ärger, Angst, Trauer, Unruhe, Zuneigung	0.70	0.23	0.18	0.02	-0.28
10	Schüler, 17 Jahre	Gewonnenes Fußballspiel	Freude, Zuneigung, Stolz, Überraschung	0.30	0.11	0.82	0.19	-0.14
11	Angestellte, 31 Jahre	Ausbleibender Anruf	Ärger, Angst, Unruhe, Abneigung, Zuneigung	0.29	0.13	0.46	0.66	0.15
12	Polizist, 24 Jahre	Frisch verheiratet	Unruhe, Freude, Stolz, Zuneigung, Angst, Überraschung	0.30	0.34	0.71	-0.19	0.12
Erklärte Varianz in % (pro Komponente)				24.56	11.88	9.85	8.66	8.35

Anmerkungen:

Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Promax mit Kaiser-Normalisierung.

(A)= Die Person ist Bewohnerin der Altenhilfeeinrichtung; (R)= Die Person ist Patientin in der Rehaklinik.

¹ **Fettgedruckt:** Emotionen, die als „stark vorhanden“ eingestuft wurden; Normal gedruckt: Emotionen, die als „mittel vorhanden“ eingestuft wurden.² Aus Gründen der Übersichtlichkeit sind Ladungen >0.30 **fettgedruckt**.*Tabelle A-17: Komponentenkorrelationsmatrix*

Komponente	1	2	3	4	5
1					
2	0.31				
3	0.45	0.24			
4	0.08	0.05	0.09		
5	-0.05	0.08	0.04	-0.18	

Anmerkungen:

Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Promax mit Kaiser-Normalisierung.

TEMINT, Stichprobe 1, Scoring Konsens-AnteilDateneignung: KMO=0.60; Bartlett-Test signifikant ($p=0.00$)*Tabelle A-18: Strukturmatrix einer Hauptkomponentenanalyse der Situationen des TEMINT, Stichprobe 1, Konsens-Anteil*

Sit.	Person	Ereignis	Dominante Emotion ¹	Komponente ²			
				1	2	3	4
1	Informatikerin, 30 Jahre	Katze vergiftet	Unruhe, Zuneigung, Trauer, Angst, Überraschung, Schuld	0.04	0.16	0.22	0.53
2	Zahnarzthelferin, 23 Jahre	Geburt	Angst, Unruhe, Stolz, Freude, Zuneigung	0.15	0.72	-0.03	0.00
3	Student, 30 Jahre	Computerproblem	Abneigung, Ärger, Angst, Unruhe, Schuld, Trauer	0.64	-0.19	0.34	-0.08
4	Beamten, 22 Jahre	Streit mit Kollegin	Abneigung, Unruhe, Ärger, Stolz, Überraschung, Trauer, Angst	0.70	-0.13	0.29	0.37
5	Sachbearbeiter, 29 Jahre	Zufälliges Treffen	Freude, Überraschung	0.70	0.08	-0.01	-0.28
6	Lagerist, 34 Jahre	Unfreundlicher Kunde	Abneigung, Ärger, Unruhe, Angst	0.56	-0.07	0.29	0.57
7	Diplom-Ingenieur, 69 Jahre	Zeitungsmeldung	Abneigung, Ärger, Unruhe	0.21	0.00	0.75	0.10
8	Lehrerin, 58 Jahre	Erkrankte Schwägerin	Angst, Unruhe, Trauer, Zuneigung, Überraschung	-0.04	0.68	0.07	0.22
9	Beamter, 44 Jahre	Eheprobleme	Ärger, Angst, Trauer, Unruhe, Zuneigung	-0.13	0.69	0.10	0.11
10	Schüler, 17 Jahre	Gewonnenes Fußballspiel	Freude, Zuneigung, Stolz, Überraschung	-0.11	0.10	-0.28	0.59
11	Angestellte, 31 Jahre	Ausbleibender Anruf	Ärger, Angst, Unruhe, Abneigung, Zuneigung	0.09	0.12	0.72	0.06
12	Polizist, 24 Jahre	Frisch verheiratet	Unruhe, Freude, Stolz, Zuneigung, Angst, Überraschung	0.49	0.23	-0.28	0.15
Erklärte Varianz in % (pro Komponente)				18.43	14.43	10.87	9.03

Anmerkungen:

Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Promax mit Kaiser-Normalisierung.

(A)= Die Person ist Bewohnerin der Altenhilfeeinrichtung; (R)= Die Person ist Patientin in der Rehaklinik.

¹ **Fettgedruckt:** Emotionen, die als „stark vorhanden“ eingestuft wurden; Normal gedruckt: Emotionen, die als „mittel vorhanden“ eingestuft wurden.² Aus Gründen der Übersichtlichkeit sind Ladungen >0.30 **fettgedruckt**.*Tabelle A-19: Komponentenkorrelationsmatrix*

Komponente	1	2	3	4
1				
2	-0.03			
3	0.18	-0.06		
4	0.11	0.06	0.12	

Anmerkungen:

Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Promax mit Kaiser-Normalisierung.

TEMINT, Stichprobe 2a, Scoring Target-ÜbereinstimmungDateneignung: KMO=0.67; Bartlett-Test signifikant ($p=0.00$)*Tabelle A-20: Strukturmatrix einer Hauptkomponentenanalyse der Situationen des TEMINT, Stichprobe 2a, Target-Übereinstimmung*

Sit.	Person	Ereignis	Dominante Emotion ¹	Komponente ²				
				1	2	3	4	5
1	Informatikerin, 30 Jahre	Katze vergiftet	Unruhe, Zuneigung, Trauer, Angst, Überraschung, Schuld	0.73	0.12	0.22	0.20	0.12
2	Zahnarzthelferin, 23 Jahre	Geburt	Angst, Unruhe, Stolz, Freude, Zuneigung	0.24	0.03	-0.02	0.79	-0.05
3	Student, 30 Jahre	Computerproblem	Abneigung, Ärger, Angst, Unruhe, Schuld, Trauer	0.51	0.16	0.56	-0.10	0.10
4	Beamten, 22 Jahre	Streit mit Kollegin	Abneigung, Unruhe, Ärger, Stolz, Überraschung, Trauer, Angst	0.33	0.52	0.20	0.05	0.09
5	Sachbearbeiter, 29 Jahre	Zufälliges Treffen	Freude, Überraschung	0.04	-0.05	-0.07	-0.05	0.87
6	Lagerist, 34 Jahre	Unfreundlicher Kunde	Abneigung, Ärger, Unruhe, Angst	0.11	0.54	0.26	0.60	0.40
7	Diplom-Ingenieur, 69 Jahre	Zeitungsmeldung	Abneigung, Ärger, Unruhe	0.04	0.07	0.71	0.03	-0.01
8	Lehrerin, 58 Jahre	Erkrankte Schwägerin	Angst, Unruhe, Trauer, Zuneigung, Überraschung	0.58	0.45	-0.04	0.04	0.13
9	Beamter, 44 Jahre	Eheprobleme	Ärger, Angst, Trauer, Unruhe, Zuneigung	0.74	0.18	-0.02	0.15	0.04
10	Schüler, 17 Jahre	Gewonnenes Fußballspiel	Freude, Zuneigung, Stolz, Überraschung	0.04	0.77	0.00	0.07	-0.15
11	Angestellte, 31 Jahre	Ausbleibender Anruf	Ärger, Angst, Unruhe, Abneigung, Zuneigung	-0.31	0.21	0.65	0.53	0.00
12	Polizist, 24 Jahre	Frisch verheiratet	Unruhe, Freude, Stolz, Zuneigung, Angst, Überraschung	0.10	0.52	0.24	0.22	0.43
Erklärte Varianz in % (pro Komponente)				21.03	11.24	9.68	9.16	9.00

Anmerkungen:

Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Promax mit Kaiser-Normalisierung.

(A)= Die Person ist Bewohnerin der Altenhilfeeinrichtung; (R)= Die Person ist Patientin in der Rehaklinik.

¹ **Fettgedruckt:** Emotionen, die als „stark vorhanden“ eingestuft wurden; Normal gedruckt: Emotionen, die als „mittel vorhanden“ eingestuft wurden.² Aus Gründen der Übersichtlichkeit sind Ladungen >0.30 **fettgedruckt**.*Tabelle A-21: Komponentenkorrelationsmatrix*

Komponente	1	2	3	4	5
1					
2	0.16				
3	0.02	0.22			
4	-0.03	0.25	0.23		
5	0.10	0.18	0.13	0.13	

Anmerkungen:

Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Promax mit Kaiser-Normalisierung.

TEMINT, Stichprobe 2a, Scoring Target-TrefferDateneignung: KMO=0.69; Bartlett-Test signifikant ($p=0.00$)*Tabelle A-22: Strukturmatrix einer Hauptkomponentenanalyse der Situationen des TEMINT, Stichprobe 2a, Target-Treffer*

Sit.	Person	Ereignis	Dominante Emotion ¹	Komponente ²				
				1	2	3	4	5
1	Informatikerin, 30 Jahre	Katze vergiftet	Unruhe, Zuneigung, Trauer, Angst, Überraschung, Schuld	0.72	0.11	0.27	0.11	0.17
2	Zahnarzthelferin, 23 Jahre	Geburt	Angst, Unruhe, Stolz, Freude, Zuneigung	0.20	0.12	0.83	0.04	-0.02
3	Student, 30 Jahre	Computerproblem	Abneigung, Ärger, Angst, Unruhe, Schuld, Trauer	0.66	0.10	-0.24	0.32	0.02
4	Beamten, 22 Jahre	Streit mit Kollegin	Abneigung, Unruhe, Ärger, Stolz, Überraschung, Trauer, Angst	0.48	0.26	0.12	0.13	-0.18
5	Sachbearbeiter, 29 Jahre	Zufälliges Treffen	Freude, Überraschung	0.06	0.09	-0.03	-0.09	0.83
6	Lagerist, 34 Jahre	Unfreundlicher Kunde	Abneigung, Ärger, Unruhe, Angst	0.23	0.64	0.49	0.11	0.06
7	Diplom-Ingenieur, 69 Jahre	Zeitungsmeldung	Abneigung, Ärger, Unruhe	0.28	0.06	-0.17	0.68	0.05
8	Lehrerin, 58 Jahre	Erkrankte Schwägerin	Angst, Unruhe, Trauer, Zuneigung, Überraschung	0.64	0.35	0.10	-0.20	-0.14
9	Beamter, 44 Jahre	Eheprobleme	Ärger, Angst, Trauer, Unruhe, Zuneigung	0.76	0.13	0.10	0.00	0.04
10	Schüler, 17 Jahre	Gewonnenes Fußballspiel	Freude, Zuneigung, Stolz, Überraschung	0.20	0.61	-0.08	-0.26	-0.60
11	Angestellte, 31 Jahre	Ausbleibender Anruf	Ärger, Angst, Unruhe, Abneigung, Zuneigung	-0.17	0.09	0.39	0.70	-0.14
12	Polizist, 24 Jahre	Frisch verheiratet	Unruhe, Freude, Stolz, Zuneigung, Angst, Überraschung	0.16	0.75	0.00	0.07	0.02
Erklärte Varianz in % (pro Komponente)				22.31	10.91	9.69	9.42	8.59

Anmerkungen:

Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Promax mit Kaiser-Normalisierung.

(A)= Die Person ist Bewohnerin der Altenhilfeeinrichtung; (R)= Die Person ist Patientin in der Rehaklinik.

¹ **Fettgedruckt:** Emotionen, die als „stark vorhanden“ eingestuft wurden; Normal gedruckt: Emotionen, die als „mittel vorhanden“ eingestuft wurden.² Aus Gründen der Übersichtlichkeit sind Ladungen >0.30 **fettgedruckt**.*Tabelle A-23: Komponentenkorrelationsmatrix*

Komponente	1	2	3	4	5
1					
2	0.26				
3	0.01	0.14			
4	0.05	-0.07	0.09		
5	0.00	-0.20	0.02	0.09	

Anmerkungen:

Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Promax mit Kaiser-Normalisierung.

Anhang F. Formelsammlung

Prüfung der Unterschiedlichkeit zweier Korrelationen

Für die Prüfung der Unterschiedlichkeit **zweier Korrelationen** r_{ab} und r_{ac} **innerhalb einer Stichprobe** unter Berücksichtigung der Korrelation r_{bc} nach der Formel aus Bortz (2005, S. 222-223).

$$H_0: \rho_{ab} = \rho_{ac}.$$

Fishers-Z-Transformation der Korrelationen nach folgender Formel:

$$r'_{ab} = \frac{1}{2} * \ln\left(\frac{1+r_{ab}}{1-r_{ab}}\right) \text{ und } r'_{ac} = \frac{1}{2} * \ln\left(\frac{1+r_{ac}}{1-r_{ac}}\right) \quad (\text{Formel 8})$$

\ln = Logarithmus zur Basis e

Durch die Transformation nach Fisher nähern sich die Korrelationskoeffizienten einer zwei-dimensionalen Normalverteilung an (Sachs & Hedderich, 2006, S. 548-549).

Berechnung der **Prüfgröße z**:

$$z = \frac{\sqrt{(n-3)} * (r'_{ab} - r'_{ac})}{\sqrt{(2-2 * CV_1)}} \quad (\text{Formel 9})$$

n = Stichprobenumfang

Schätzung der Kovarianz CV_1 der Korrelationsverteilungen von r_{ab} und r_{ac} :

$$CV_1 = \frac{1}{(1-r_a^2)^2} * (r_{bc} * (1 - 2 * r_a^2) - 0.5 * r_a^2 * (1 - 2 * r_a^2 - r_{bc}^2)) \quad (\text{Formel 10})$$

$$r_a = (r_{ab} + r_{ac})/2 \quad (\text{Formel 11})$$

Prüfung der Unterschiedlichkeit **zweier Korrelationskoeffizienten** r_{ab} und r_{cd} **in zwei Stichproben** nach der Formel aus Bortz (2005, S. 222-223).

$$H_0: \rho_{ab} = \rho_{cd}.$$

Fishers-Z-Transformation der Korrelationen nach folgender Formel:

$$r'_{ab} = \frac{1}{2} * \ln\left(\frac{1+r_{ab}}{1-r_{ab}}\right) \text{ und } r'_{cd} = \frac{1}{2} * \ln\left(\frac{1+r_{cd}}{1-r_{cd}}\right) \quad (\text{Formel 12})$$

\ln = Logarithmus zur Basis e

Berechnung der **Prüfgröße z**:

$$z = \frac{r'_{ab} - r'_{cd}}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}} \quad (\text{Formel 13})$$

Berechnung der Effektstärke

Berechnung der Effektstärke d für abhängige Stichproben nach Bortz und Döring (2006, S. 609), bei unterschiedlicher Varianz

$$d = \frac{\bar{x}_a - \bar{x}_b}{\hat{s}} \quad (\text{Formel 14})$$

$$\hat{s} = \sqrt{\frac{\hat{s}_a^2 + \hat{s}_b^2}{2}} \quad (\text{gepoolte Standardabweichung}) \quad (\text{Formel 15})$$

Berechnung der Effektstärke d für unabhängige Stichproben nach Bortz und Döring (2006, S. 607), bei heterogenen Varianzen

$$d = \frac{\bar{x}_a - \bar{x}_b}{\hat{s}} \quad (\text{Formel 16})$$

$$\hat{s} = \sqrt{\frac{(n_a - 1) * \hat{s}_a^2 + (n_b - 1) * \hat{s}_b^2}{(n_a - 1) + (n_b - 1)}} \quad (\text{gepoolte Standardabweichung}) \quad (\text{Formel 17})$$

Anhang G. Antwortbias auf Seiten der Probanden*Tabelle A-24: Korrelation der Kennwerte zum Antwortverhalten mit den Gesamtestwerten bei unterschiedlichen Scorings (ViTEIP-V2 bzw. TEMINT)*

Stich- probe	Test	Kennwert „Häufigkeit Antwortoption...	Gesamtestwert ermittelt anhand Scoring:			
			Target- Übereinst.	Target- Treffer	Konsens- Anteil	Konsens- Modal
1	ViTEIP-V2	...Emotion nicht vorhanden“	0.68**	0.72**	0.57**	0.63**
		...Emotion schwach vorhanden“	0.11	-0.25**	-0.17	-0.30**
		...Emotion mittel vorhanden“	-0.50**	-0.55**	-0.47**	-0.39**
		...Emotion stark vorhanden“	-0.55**	-0.27**	-0.22*	-0.25**
	TEMINT	...Emotion nicht oder sehr schwach vorhanden“	-0.77**	-0.63**	-0.07	-0.24*
		...Emotion schwach bis mittel vorhanden“	-0.31**	-0.50**	-0.56**	-0.44**
		...Emotion stark bis sehr stark vorhanden“	0.90**	0.94**	0.52**	0.56**
2a	ViTEIP-V2	...Emotion nicht vorhanden“	0.51**	0.68**	0.09	0.26**
		...Emotion schwach vorhanden“	0.52**	0.00	-0.42**	-0.35**
		...Emotion mittel vorhanden“	-0.05	-0.42**	-0.15	-0.26**
		...Emotion stark vorhanden“	-0.77**	-0.37**	0.33**	0.17*
	TEMINT	...Emotion nicht oder sehr schwach vorhanden“	-0.65**	-0.45**	0.07	-0.12
		...Emotion schwach bis mittel vorhanden“	-0.31**	-0.58**	-0.45**	-0.46**
		...Emotion stark bis sehr stark vorhanden“	0.76**	0.87**	0.38**	0.52**
2b	ViTEIP-V2	...Emotion nicht vorhanden“	0.47**	0.58**	0.83**	0.87**
		...Emotion schwach vorhanden“	0.16	-0.18*	-0.43**	-0.49**
		...Emotion mittel vorhanden“	-0.18*	-0.37**	-0.68**	-0.71**
		...Emotion stark vorhanden“	-0.60**	-0.47**	-0.51**	-0.51**
	TEMINT	...Emotion nicht oder sehr schwach vorhanden“	-0.78**	-0.66**	-0.23*	-0.13
		...Emotion schwach bis mittel vorhanden“	0.04	-0.26**	-0.30**	-0.37**
		...Emotion stark bis sehr stark vorhanden“	0.81**	0.90**	0.55**	0.50**

*Anmerkungen:*Zweiseitige Testung, * $p < 0.05$. ** $p < 0.01$.

Für eine detaillierte Beschreibung der Scorings siehe Kapitel 3.3.1.

Häufigkeit Antwortoption= Häufigkeit, mit der die Probanden im ViTEIP bzw. TEMINT die jeweiligen Antwortoptionen gewählt haben.

Tabelle A-25: Korrelation der Häufigkeiten der gewählten Antwortoptionen bei ViTEIP-V2 und TEMINT

		TEMINT Häufigkeit Antwortoption		
		„Emotion nicht oder sehr schwach vorhanden“	„Emotion schwach bis mittel vorhanden“	„Emotion stark bis sehr stark vorhanden“
Stichprobe 1 (N=113)				
ViTEIP	„Emotion nicht vorhanden“	0.53**	-0.27**	-0.23**
Häufigkeit	„Emotion schwach vorhanden“	-0.02	0.35**	-0.27**
Antwort-	„Emotion mittel vorhanden“	-0.41**	0.22*	0.16*
option	„Emotion stark vorhanden“	-0.32**	-0.14	0.38**
Stichprobe 2a (N=138)				
ViTEIP	„Emotion nicht vorhanden“	0.48**	-0.12	-0.24**
Häufigkeit	„Emotion schwach vorhanden“	-0.08	0.49**	-0.39**
Antwort-	„Emotion mittel vorhanden“	-0.32**	0.36**	-0.11
option	„Emotion stark vorhanden“	-0.22**	-0.48**	0.61**
Stichprobe 2b (N=122)				
ViTEIP	„Emotion nicht vorhanden“	0.42**	-0.19*	-0.25**
Häufigkeit	„Emotion schwach vorhanden“	-0.28**	0.38**	-0.05
Antwort-	„Emotion mittel vorhanden“	-0.38**	0.35**	0.07
option	„Emotion stark vorhanden“	-0.20*	-0.17*	0.33**

Anmerkungen:

Zweiseitige Testung, * $p < 0.05$. ** $p < 0.01$.**fettgedruckt** hinterlegt sind die Korrelationen bei vergleichbaren Antwortkategorien der beiden Tests.

Häufigkeit Antwortoption= Häufigkeit, mit der die Probanden im ViTEIP bzw. TEMINT die jeweiligen Antwortoptionen gewählt haben.

Anhang H. Übertragung der Signalentdeckungstheorie auf Emotionale-Intelligenz-Tests

Tabelle A-26: Korrelation des ViTEIP-V2 und TEMINT mit anderen Verfahren (Stichprobe 2a)

Kennwert	ViTEIP			TEMINT		
	\hat{d}'	eSens	eSpez	\hat{d}'	eSens	eSpez
d2-R_ Konzentrationsleistung	0.05	0.08	0.02	-0.20*	0.12	-0.12
Merkfähigkeit	-0.05	0.13	-0.18*	-0.08	0.24**	-0.12
Empathie	0.06	0.03	-0.01	-0.01	-0.01	-0.03
EKF_Gesamt	-0.14	-0.08	-0.02	-0.09	-0.09	-0.10
EKF_EE	-0.17*	-0.03	0.03	-0.13	-0.05	-0.17*
EKF_EA	-0.03	-0.15*	-0.01	0.00	-0.09	-0.03
EKF_RG	-0.04	0.04	-0.08	-0.08	-0.17*	-0.03
EKF_EX	-0.15*	-0.08	0.00	-0.04	0.06	-0.05
Geschlecht ¹	0.04	-0.04	0.04	0.00	-0.02	0.02
Alter ²	-0.06	-0.06	-0.07	-0.10	-0.02	-0.09
Schulabschluss	0.08	0.05	0.07	-0.02	-0.06	0.03
Ausbildungsjahr	0.02	0.02	-0.02	0.01	-0.10	0.06
Erfahrungen im Pflegebereich ²	-0.08	0.10	-0.07	0.08	-0.06	0.12

Anmerkungen:

Korrelation nach Pearson. Einseitige Testung, *p<0.05. **p<0.01.

\hat{d}' = Detektionsleistung, d2-R= Aufmerksamkeits-Konzentrations-Test d2, revidierte Form (Brickenkamp, Schmidt-Atzert, & Liepmann, 2010), Kennwert Konzentrationsleistung. Merkfähigkeit= WIT Skala Merkfähigkeit (Kersting, Althoff, & Jäger, 2008). Empathie= Situationsbezogener Empathiefragebogen (Holz-Ebeling & Steinmetz, 1994). EKF_EE= EKF-Skala Erkennen eigener Emotionen; EKF_EA= EKF-Skala Erkennen Emotionen anderer; EKF_RG= EKF-Skala Regulation und Kontrolle; EKF_EX= EKF-Skala Emotionale Expressivität.

¹ Kodierung: 0= männlich, 1= weiblich.

² in Jahren.

Tabelle A-27: Korrelation des ViTEIP-V2 und TEMINT mit anderen Verfahren (Stichprobe 2b)

Kennwert	ViTEIP			TEMINT		
	\hat{d}'	eSens	eSpez	\hat{d}'	eSens	eSpez
d2-R_ Konzentrationsleistung	0.17*	-0.01	0.07	0.08	0.12	-0.14
Merkfähigkeit	0.07	-0.04	0.10	0.06	0.09	-0.10
Empathie	-0.01	-0.01	-0.05	0.16*	0.15	-0.06
EKF_Gesamt	-0.07	-0.09	-0.04	-0.13	0.06	-0.22**
EKF_EE	-0.01	0.00	0.04	-0.19*	0.06	-0.19*
EKF_EA	-0.06	-0.05	-0.05	-0.06	0.11	-0.15
EKF_RG	-0.10	-0.07	-0.06	-0.07	0.10	-0.13
EKF_EX	-0.07	-0.16*	-0.07	-0.05	-0.08	-0.17*
Geschlecht ¹	-0.06	0.11	-0.14	0.19*	0.23**	0.05
Alter ²	0.18*	-0.08	0.15*	-0.07	-0.20*	-0.06
Schulabschluss	-0.07	0.09	-0.06	-0.07	0.07	0.10
Ausbildungsjahr	0.03	-0.02	0.16*	-0.20*	-0.25**	-0.06
Erfahrungen im Pflegebereich ²	0.06	-0.02	0.06	-0.08	-0.16*	-0.07

Anmerkungen:

Korrelation nach Pearson. Einseitige Testung, *p<0.05. **p<0.01.

\hat{d}' = Detektionsleistung, d2-R= Aufmerksamkeits-Konzentrations-Test d2, revidierte Form (Brickenkamp, Schmidt-Atzert, & Liepmann, 2010), Kennwert Konzentrationsleistung. Merkfähigkeit= WIT Skala Merkfähigkeit (Kersting, Althoff, & Jäger, 2008). Empathie= Situationsbezogener Empathiefragebogen (Holz-Ebeling & Steinmetz, 1994). EKF_EE= EKF-Skala Erkennen eigener Emotionen; EKF_EA= EKF-Skala Erkennen Emotionen anderer; EKF_RG= EKF-Skala Regulation und Kontrolle; EKF_EX= EKF-Skala Emotionale Expressivität.

¹ Kodierung: 0= männlich, 1= weiblich.

² in Jahren.

Tabelle A-28: Korrelationsmatrix SET-Kennwerte des ViTEIP-V2 und TEMINT (Stichprobe 1)

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.
1. ViTEIP-V2 Target-Übereinstimmung																			
2. ViTEIP-V2 Konsens-Anteil	0.60**																		
3. ViTEIP-V2 Detektionsleistung \hat{d}'	0.73**	0.37**																	
4. ViTEIP-V2 Reaktionsneigung \hat{c}	0.68**	0.60**	0.59**																
5. ViTEIP-V2 Identifikation	-0.34**	-0.45**	-0.05	-0.81**															
6. ViTEIP-V2 Verpasser	0.34**	0.45**	0.05	0.81**	-1.00**														
7. ViTEIP-V2 korrekte Zurückweisung	0.82**	0.57**	0.71**	0.90**	-0.65**	0.65**													
8. ViTEIP-V2 falscher Alarm	-0.82**	-0.57**	-0.71**	-0.90**	0.65**	-0.65**	-1.00**												
9. ViTEIP-V2 emotionale Sensitivität	0.15	-0.07	0.15	-0.47**	0.72**	-0.72**	-0.33**	0.33**											
10. ViTEIP-V2 emotionale Spezifität	0.88**	0.60**	0.63**	0.88**	-0.67**	0.67**	0.93**	-0.93**	-0.33**										
11. TEMINT Target-Übereinstimmung	-0.26**	-0.13	-0.13	-0.32**	0.32**	-0.32**	-0.30**	0.30**	0.25**	-0.37**									
12. TEMINT Konsens-Anteil	0.17	0.43**	0.09	0.21*	-0.19*	0.19*	0.20*	-0.20*	0.00	0.17	0.59**								
13. TEMINT Detektionsleistung \hat{d}'	0.14	0.23*	0.05	0.06	-0.07	0.07	0.12	-0.12	-0.01	0.14	0.36**	0.66**							
14. TEMINT Reaktionsneigung \hat{c}	0.39**	0.47**	0.22*	0.42**	-0.41**	0.41**	0.45**	-0.45**	-0.21*	0.47**	-0.26**	0.44**	0.74**						
15. TEMINT Identifikation	-0.36**	-0.35**	-0.22*	-0.51**	0.49**	-0.49**	-0.47**	0.47**	0.30**	-0.49**	0.85**	0.25**	0.20*	-0.50**					
16. TEMINT Verpasser	0.36**	0.35**	0.22*	0.51**	-0.49**	0.49**	0.47**	-0.47**	-0.30**	0.49**	-0.85**	-0.25**	-0.20*	0.50**	-1.00**				
17. TEMINT korrekte Zurückweisung	0.28**	0.40**	0.22*	0.31**	-0.25**	0.25**	0.31**	-0.31**	-0.12	0.32**	0.03	0.63**	0.79**	0.85**	-0.20*	0.20*			
18. TEMINT falscher Alarm	-0.28**	-0.40**	-0.22*	-0.31**	0.25**	-0.25**	-0.31**	0.31**	0.12	-0.32**	-0.03	-0.63**	-0.79**	-0.85**	0.20*	-0.20*	-1.00**		
19. TEMINT emotionale Sensitivität	-0.30**	-0.19*	-0.17	-0.38**	0.36**	-0.36**	-0.35**	0.35**	0.27**	-0.42**	0.99**	0.48**	0.23*	-0.39**	0.88**	-0.88**	-0.12	0.12	
20. TEMINT emotionale Spezifität	0.29**	0.38**	0.23*	0.33**	-0.27**	0.27**	0.31**	-0.31**	-0.13	0.33**	0.04	0.61**	0.78**	0.83**	-0.19*	0.19*	0.96**	-0.96**	-0.12

Anmerkungen:

Korrelation nach Pearson. Einseitige Testung, * $p < 0.05$, ** $p < 0.01$.

Für eine detaillierte Beschreibung der Scorings siehe Kapitel 3.3.1 bzw. 4.2

Tabelle A-29: Korrelationsmatrix SET-Kennwerte des ViTEIP-V2 und TEMINT (Stichprobe 2a)

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.
1. ViTEIP-V2 Target-Übereinstimmung																			
2. ViTEIP-V2 Konsens-Anteil	0.06																		
3. ViTEIP-V2 Detektionsleistung \hat{d}'	0.45**	0.06																	
4. ViTEIP-V2 Reaktionsneigung \hat{c}	0.47**	0.18*	0.19*																
5. ViTEIP-V2 Identifikation	-0.32**	-0.06	0.08	-0.87**															
6. ViTEIP-V2 Verpasser	0.32**	0.06	-0.08	0.87**	-1.00**														
7. ViTEIP-V2 korrekte Zurückweisung	0.60**	0.11	0.51**	0.90**	-0.78**	0.77**													
8. ViTEIP-V2 falscher Alarm	-0.60**	-0.11	-0.51**	-0.90**	0.78**	-0.77**	-1.00**												
9. ViTEIP-V2 emotionale Sensitivität	0.23**	0.14	0.07	-0.58**	0.70**	-0.70**	-0.53**	0.53**											
10. ViTEIP-V2 emotionale Spezifität	0.87**	-0.02	0.41**	0.77**	-0.67**	0.67**	0.86**	-0.86**	-0.29**										
11. TEMINT Target-Übereinstimmung	-0.33**	0.16	-0.13	-0.42**	0.37**	-0.37**	-0.42**	0.42**	0.26**	-0.45**									
12. TEMINT Konsens-Anteil	-0.02	0.33**	-0.01	-0.06	0.09	-0.09	-0.08	0.08	0.20*	-0.12	0.58**								
13. TEMINT Detektionsleistung \hat{d}'	0.07	0.10	0.05	0.02	0.03	-0.03	0.00	0.00	0.07	0.04	0.34**	0.68**							
14. TEMINT Reaktionsneigung \hat{c}	0.22**	0.11	0.11	0.35**	-0.28**	0.28**	0.33**	-0.33**	-0.15	0.30**	-0.22*	0.47**	0.77**						
15. TEMINT Identifikation	-0.23**	-0.02	-0.14	-0.50**	0.47**	-0.47**	-0.51**	0.51**	0.36**	-0.41**	0.78**	0.16	0.09	-0.55**					
16. TEMINT Verpasser	0.23**	0.02	0.14	0.50**	-0.47**	0.47**	0.51**	-0.51**	-0.36**	0.41**	-0.78**	-0.16	-0.09	0.55**	-1.00**				
17. TEMINT korrekte Zurückweisung	0.23**	0.21*	0.14	0.23**	-0.16	0.16	0.23**	-0.23**	0.02	0.22**	0.03	0.65**	0.79**	0.84**	-0.29**	0.29**			
18. TEMINT falscher Alarm	-0.23**	-0.21*	-0.14	-0.23**	0.16	-0.16	-0.23**	0.23**	-0.02	-0.22**	-0.03	-0.65**	-0.79**	-0.84**	0.29**	-0.29**	-1.00**		
19. TEMINT emotionale Sensitivität	-0.39**	0.11	-0.17	-0.45**	0.38**	-0.38**	-0.45**	0.45**	0.23**	-0.50**	0.97**	0.42**	0.14	-0.41**	0.82**	-0.82**	-0.21*	0.21*	
20. TEMINT emotionale Spezifität	0.26**	0.21*	0.14	0.16	-0.08	0.08	0.15	-0.15	0.11	0.20*	0.06	0.59**	0.76**	0.76**	-0.21*	0.21*	0.92**	-0.92**	-0.19*

Anmerkungen:

Korrelation nach Pearson. Einseitige Testung, * $p < 0.05$. ** $p < 0.01$.

Für eine detaillierte Beschreibung der Scorings siehe Kapitel 3.3.1 bzw. 4.2

Tabelle A-30: Korrelationsmatrix SET-Kennwerte des ViTEIP-V2 und TEMINT (Stichprobe 2b)

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.
1. ViTEIP-V2 Target-Übereinstimmung																			
2. ViTEIP-V2 Konsens-Anteil	0.57**																		
3. ViTEIP-V2 Detektionsleistung \hat{d}'	0.82**	0.49**																	
4. ViTEIP-V2 Reaktionsneigung \hat{c}	0.53**	0.85**	0.46**																
5. ViTEIP-V2 Identifikation	-0.16	-0.70**	0.00	-0.88**															
6. ViTEIP-V2 Verpasser	0.16	0.70**	-0.00	0.88**	-1.00**														
7. ViTEIP-V2 korrekte Zurückweisung	0.69**	0.85**	0.70**	0.94**	-0.71**	0.71**													
8. ViTEIP-V2 falscher Alarm	-0.69**	-0.85**	-0.70**	-0.94**	0.71**	-0.71**	-1.00**												
9. ViTEIP-V2 emotionale Sensitivität	0.22*	-0.41**	0.14	-0.66**	0.82**	-0.82**	-0.48**	0.48**											
10. ViTEIP-V2 emotionale Spezifität	0.81**	0.78**	0.68**	0.89**	-0.64**	0.64**	0.94**	-0.94**	-0.40**										
11. TEMINT Target-Übereinstimmung	-0.06	-0.20*	-0.05	-0.34**	0.37**	-0.37**	-0.30**	0.30**	0.36**	-0.27**									
12. TEMINT Konsens-Anteil	0.12	0.13	0.09	0.01	0.05	-0.05	0.03	-0.03	0.11	0.04	0.64**								
13. TEMINT Detektionsleistung \hat{d}'	0.03	0.08	0.07	0.08	-0.05	0.05	0.08	-0.08	-0.03	0.049	0.37**	0.64**							
14. TEMINT Reaktionsneigung \hat{c}	0.08	0.29**	0.10	0.35**	-0.35**	0.35**	0.32**	-0.32**	-0.31**	0.26**	-0.37**	0.27**	0.64**						
15. TEMINT Identifikation	-0.08	-0.31**	-0.05	-0.39**	0.43**	-0.43**	-0.34**	0.34**	0.38**	-0.30**	0.87**	0.38**	0.22*	-0.59**					
16. TEMINT Verpasser	0.08	0.31**	0.05	0.39**	-0.43**	0.43**	0.34**	-0.34**	-0.38**	0.30**	-0.87**	-0.38**	-0.22*	0.59**	-1.00**				
17. TEMINT korrekte Zurückweisung	0.09	0.23*	0.07	0.25**	-0.25**	0.25**	0.23*	-0.23*	-0.19*	0.20*	0.03	0.54**	0.81**	0.79**	-0.17	0.17			
18. TEMINT falscher Alarm	-0.09	-0.23*	-0.07	-0.25**	0.25**	-0.25**	-0.23*	0.23*	0.19*	-0.20*	-0.03	-0.54**	-0.81**	-0.79**	0.17	-0.17	-1.00**		
19. TEMINT emotionale Sensitivität	-0.08	-0.24**	-0.06	-0.38**	0.41**	-0.41**	-0.34**	0.34**	0.38**	-0.31**	0.99**	0.57**	0.27**	-0.47**	0.90**	-0.90**	-0.10	0.10	
20. TEMINT emotionale Spezifität	0.09	0.15	0.03	0.14	-0.13	0.13	0.13	-0.13	-0.08	0.18*	0.29**	0.53**	0.66**	0.52**	0.03	-0.03	0.76**	-0.76**	0.12

Anmerkungen:

Korrelation nach Pearson. Einseitige Testung, * $p < 0.05$. ** $p < 0.01$.

Für eine detaillierte Beschreibung der Scorings siehe Kapitel 3.3.1 bzw. 4.2

Anhang I. Vergleich der Antwortformate

Tabelle A-31: Akzeptanzwerte des ViTEIP-V2 bei Antwortformat „Ankreuzen“ und „Zuordnen“

Kriterien ¹	„Ankreuzen“		„Zuordnen“		Mittelwertvergleich: „Ankreuzen“ und „Zuordnen“	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>d</i>
Nützlichkeit	1.72	0.81	1.81	0.80	0.96	0.01
Spas	1.76	0.80	1.98	0.88	0.10	0.02
Akzeptanz	1.58	0.82	1.88	1.00	0.68	0.03
Fairness	1.42	0.84	1.74	0.95	0.72	0.03

Anmerkungen:

„Ankreuzen“ bzw. „Zuordnen“= Antwortformate des ViTEIP, Beschreibung siehe Kapitel 1.5.3

M= Mittelwert; *SD*= Standardabweichung. *t*= T-Test bei gepaarten Stichproben, *df*=112. *d*= Effektstärke.

¹ Beantwortung auf einer vierstufigen Skala mit „0=stimme nicht zu“, „1=stimme eher nicht zu“, „2=stimme eher zu“, „3=stimme zu“.

Tabelle A-32: Korrelation der Kennwerte des ViTEIP-V2 bei unterschiedlichen Antwortformaten und verschiedenen Scorings zu anderen Variablen

	Target-Übereinstimmung			Target-Treffer			Konsens-Anteil			Konsens-Modal		
	„An- kreuzen“	„Zu- ordnen“	z	„An- kreuzen“	„Zu- ordnen“	z	„An- kreuzen“	„Zu- ordnen“	z	„An- kreuzen“	„Zu- ordnen“	z
d2-R_KL	0.06	0.07	-0.08	0.12	0.12	0.00	0.18*	0.00	1.45	0.16*	0.04	0.97
Merkfähigkeit	-0.12	0.08	-1.60	-0.11	0.02	-1.04	-0.05	-0.03	-0.16	-0.08	0.03	-0.88
Empathie	0.00	-0.06	0.48	0.02	-0.09	0.88	-0.14	-0.12	-0.16	-0.14	-0.09	-0.40
EKF_Gesamt	-0.07	-0.11	0.32	-0.05	-0.11	0.48	-0.06	-0.04	-0.16	-0.04	-0.02	-0.16
EKF_EE	0.01	0.05	-0.32	-0.07	-0.01	-0.48	0.00	-0.01	0.08	-0.03	-0.01	-0.16
EKF_EA	-0.09	-0.09	0.00	-0.01	-0.19*	1.45	-0.08	-0.18*	0.81	0.03	-0.13	1.28
EKF_RG	-0.06	-0.11	0.40	-0.02	-0.10	0.64	-0.01	0.02	-0.24	-0.01	0.04	-0.40
EKF_EX	-0.05	-0.18*	1.05	-0.04	-0.06	0.16	-0.07	0.01	-0.64	-0.09	0.03	-0.96
Alter ¹	-0.10	0.12	-1.76	0.03	0.19*	-1.29	0.00	0.15	-1.20	0.04	0.13	-0.72
Geschlecht ²	0.02	-0.07	0.72	0.14*	-0.07	1.68	0.18*	-0.12	2.41	0.22**	-0.14	2.90
Erfahrungen ³	-0.03	0.05	-0.64	0.00	0.13	-1.04	0.04	0.16*	-0.97	-0.01	0.13	-1.12

	Profil-Gesamtwert			Detektionsleistung			Emotionale Sensitivität			Emotionale Spezifität		
	„An- kreuzen“	„Zu- ordnen“	z	„An- kreuzen“	„Zu- ordnen“	z	„An- kreuzen“	„Zu- ordnen“	z	„An- kreuzen“	„Zu- ordnen“	z
d2-R_KL	0.07	0.02	0.40	0.05	0.17*	-0.97	0.08	-0.01	0.72	0.02	0.07	-0.40
Merkfähigkeit	-0.03	0.05	-0.64	-0.05	0.07	-0.96	0.13	-0.04	1.36	-0.18*	0.10	-2.25
Empathie	-0.10	0.08	-1.44	0.06	-0.01	0.56	0.03	-0.01	0.32	-0.01	-0.05	0.32
EKF_Gesamt	-0.11	-0.06	-0.40	-0.14	-0.07	-0.56	-0.08	-0.09	0.08	-0.02	-0.04	0.16
EKF_EE	-0.01	0.04	-0.40	-0.17*	-0.01	-1.29	-0.03	0.00	-0.24	0.03	0.04	-0.08
EKF_EA	-0.11	-0.07	-0.32	-0.03	-0.06	0.24	-0.15*	-0.05	-0.80	-0.01	-0.05	0.32
EKF_RG	-0.04	-0.10	0.48	-0.04	-0.10	0.48	0.04	-0.07	0.88	-0.08	-0.06	-0.16
EKF_EX	-0.16*	-0.07	-0.73	-0.15*	-0.07	-0.64	-0.08	-0.16*	0.65	0.00	-0.07	0.56
Alter ¹	0.02	0.12	-0.80	-0.06	0.18*	-1.93	-0.06	-0.08	0.16	-0.07	0.15*	-1.76
Geschlecht ²	-0.04	-0.10	0.48	0.04	-0.06	0.80	-0.04	0.11	-1.20	0.04	-0.14	1.44
Erfahrungen ³	0.08	0.13	-0.40	-0.08	0.06	-1.12	0.10	-0.02	0.96	-0.07	0.06	-1.04

Anmerkungen:

Korrelation nach Pearson, einseitige Testung, *p<0.05, **p<0.01.

„Ankreuzen“ bzw. „Zuordnen“= Antwortformate des ViTEIP, Beschreibung siehe Kapitel 1.5.3

z= Prüfgröße zur Prüfung der Unterschiedlichkeit der Korrelationen zwischen den Antwortformaten „Ankreuzen“ und „Zuordnen“ zu den anderen Kennwerten; kritischer Wert auf dem 5%-Niveau bei zweiseitigem Test $z_{krit}=\pm 1.96$. Zur besseren Übersichtlichkeit wurden signifikante Unterschiede **fett** gedruckt.

d2-R_KL= Aufmerksamkeits-Konzentrations-Test d2, revidierte Form (Brickenkamp, Schmidt-Atzert, & Liepmann, 2010), Kennwert Konzentrationsleistung. Merkfähigkeit= WIT Skala Merkfähigkeit (Kersting, Althoff, & Jäger, 2008). Empathie= Situationsbezogener Empathiefragebogen (Holz-Ebeling & Steinmetz, 1994). EKF= Emotionale-Kompetenz-Fragebogen (Rindermann, 2007). EKF_EE= EKF-Skala Erkennen eigener Emotionen; EKF_EA= EKF-Skala Erkennen Emotionen anderer; EKF_RG= EKF-Skala Regulation und Kontrolle; EKF_EX= EKF-Skala Emotionale Expressivität.

¹ in Jahren

² Geschlecht (0=männlich, 1=weiblich)

³ Erfahrungen im Pflegebereich in Jahren.

Anhang J. Analyse der Präsentationsmodi**Tabelle A-33: Mittelwertvergleich der text- und der videobasierten Forschungsversion (ViTEIP; Scoring: Target-Übereinstimmung)**

Item	AdO	Target-Übereinstimmung				Mittelwert- vergleich Text und Video		Item	AdO	Target-Übereinstimmung				Mittelwert- vergleich Text und Video	
		Text		Video						Text		Video			
		M	S	M	S	t	d			M	S	M	S	t	d
s1e1	3	1.79	0.86	1.57	0.88	2.36*	0.25	s4e6	3	2.75	0.51	2.73	0.6	0.34	0.04
s1e2	3	2.12	0.85	2.21	0.80	-0.99	-0.11	s4e7	3	2.88	0.35	2.89	0.34	-0.21	-0.03
s1e3	0	2.74	0.59	2.58	0.69	2.33*	0.25	s4e8	1	2.27	0.69	2.15	0.7	1.88	0.17
s1e4	0	2.27	0.88	2.19	0.95	0.81	0.09	s4e9	0	2.30	0.82	2.14	0.94	2.22*	0.18
s1e5	0	1.50	0.94	1.37	0.91	1.58	0.14	s4e11	0	2.67	0.77	2.67	0.65	0.00	0.00
s1e6	2	2.27	0.60	2.25	0.45	0.25	0.04	s4e14	0	1.69	1.08	1.79	1.13	-0.97	-0.09
s1e7	1	2.04	0.73	1.81	0.71	2.55*	0.32	s6e1	0	1.99	0.97	2.17	0.97	-1.81	-0.19
s1e8	2	1.60	0.70	1.57	0.72	0.46	0.04	s6e2	0	1.12	0.92	1.05	0.92	0.77	0.08
s1e9	0	1.47	1.00	1.08	0.94	4.09*	0.40	s6e3	0	2.33	0.82	2.24	0.89	1.07	0.11
s1e12	0	2.83	0.42	2.82	0.43	0.22	0.02	s6e4	0	2.48	0.78	2.46	0.68	0.22	0.03
s1e13	0	2.93	0.29	2.86	0.37	2.03	0.21	s6e5	0	1.84	0.95	2.00	0.83	-2.07*	-0.18
s3e1	0	2.17	0.90	1.92	0.94	2.83*	0.27	s6e6	3	1.96	0.91	1.95	0.9	0.12	0.01
s3e2	0	1.88	0.90	1.78	0.95	1.04	0.11	s6e7	2	2.27	0.45	2.26	0.5	0.30	0.02
s3e4	0	1.80	0.96	1.61	0.99	2.27*	0.19	s6e9	0	2.97	0.16	2.94	0.24	1.65	0.15
s3e5	0	1.10	0.94	0.93	0.78	2.00*	0.20	s6e12	0	2.67	0.59	2.71	0.62	-0.60	-0.07
s3e6	0	2.68	0.59	2.62	0.62	0.98	0.10	s6e14	3	2.15	0.83	2.03	1.03	1.47	0.13
s3e8	0	2.80	0.58	2.96	0.19	-2.86*	-0.37	s9e1	2	2.24	0.66	2.2	0.67	0.53	0.06
s3e9	0	2.65	0.64	2.68	0.67	-0.34	-0.05	s9e2	2	2.18	0.57	2.24	0.59	-0.93	-0.10
s3e11	1	2.12	0.40	2.16	0.39	-1.04	-0.10	s9e4	0	2.39	0.82	2.33	0.84	0.80	0.07
s3e12	3	1.08	1.04	0.96	0.86	1.45	0.13	s9e5	1	1.93	0.72	2.1	0.73	-2.17*	-0.23
s3e13	1	2.26	0.55	2.25	0.51	0.14	0.02	s9e6	0	2.12	0.92	2.03	0.97	1.08	0.10
s3e14	1	2.14	0.61	2.27	0.55	-1.71	-0.22	s9e7	2	2.33	0.59	2.26	0.61	0.97	0.12
s3e15	0	2.48	0.87	2.19	0.99	2.95*	0.31	s9e8	2	2.19	0.68	2.19	0.68	-0.12	0.00
s4e1	2	2.22	0.70	2.19	0.69	0.41	0.04	s9e9	3	2.55	0.67	2.6	0.71	-0.79	-0.07
s4e2	3	2.58	0.68	2.65	0.61	-1.13	-0.11	s9e11	0	2.22	1.08	2.26	1.02	-0.44	-0.04
s4e3	0	2.71	0.62	2.77	0.5	-1.07	-0.11	s9e14	2	1.75	0.79	1.68	0.8	0.84	0.09
s4e4	0	2.00	1.00	2.00	0.96	0.00	0.00	s9e15	2	2.23	0.58	2.07	0.7	2.26*	0.25
s4e5	0	0.96	0.88	0.98	0.93	-0.19	-0.02								

Anmerkungen:

N= 113. AdO= Angabe der Originalperson.

M= Mittelwert; S= Standardabweichung. t= T-Test bei verbundenen Stichproben, df=112. d= Effektstärke.

*Auf dem Niveau 0.05 signifikanter Unterschied zwischen den Mittelwerten „Text“ und „Video“.

Tabelle A-34: Mittelwertvergleich der text- und der videobasierten Forschungsversion (ViTEIP; Scoring: Target-Treffer)

Item	AdO	Target-Treffer				Mittelwert- vergleich Text und Video		Item	AdO	Target-Treffer				Mittelwert- vergleich Text und Video	
		Text		Video						Text		Video			
		M	S	M	S	t	d			M	S	M	S	t	d
s1e1	3	0.24	0.43	0.13	0.34	2.40*	0.27	s4e6	3	0.79	0.41	0.81	0.40	-0.42	-0.04
s1e2	3	0.37	0.49	0.42	0.50	-0.87	-0.09	s4e7	3	0.89	0.31	0.90	0.30	-0.24	-0.03
s1e3	0	0.81	0.39	0.68	0.47	2.87*	0.31	s4e8	1	0.41	0.49	0.33	0.47	1.49	0.17
s1e4	0	0.50	0.50	0.50	0.50	0.00	0.00	s4e9	0	0.50	0.50	0.45	0.50	1.23	0.11
s1e5	0	0.15	0.36	0.11	0.31	1.29	0.13	s4e11	0	0.81	0.39	0.76	0.43	1.23	0.13
s1e6	2	0.35	0.48	0.26	0.44	1.48	0.19	s4e14	0	0.31	0.46	0.35	0.48	-1.00	-0.09
s1e7	1	0.27	0.45	0.17	0.38	2.09*	0.26	s6e1	0	0.39	0.49	0.50	0.50	-2.24*	-0.23
s1e8	2	0.12	0.33	0.12	0.33	0.00	0.00	s6e2	0	0.09	0.29	0.07	0.26	0.71	0.07
s1e9	0	0.15	0.36	0.08	0.27	1.91	0.22	s6e3	0	0.53	0.50	0.49	0.50	0.90	0.09
s1e12	0	0.85	0.36	0.84	0.37	0.24	0.02	s6e4	0	0.63	0.49	0.56	0.50	1.42	0.14
s1e13	0	0.94	0.24	0.87	0.34	2.17*	0.24	s6e5	0	0.30	0.46	0.32	0.47	-0.41	-0.04
s3e1	0	0.44	0.50	0.32	0.47	2.45*	0.26	s6e6	3	0.32	0.47	0.32	0.47	0.00	0.00
s3e2	0	0.28	0.45	0.27	0.45	0.18	0.02	s6e7	2	0.27	0.45	0.28	0.45	-0.16	-0.02
s3e4	0	0.28	0.45	0.22	0.42	1.47	0.14	s6e9	0	0.97	0.16	0.94	0.24	1.65	0.17
s3e5	0	0.08	0.27	0.01	0.09	2.92*	0.35	s6e12	0	0.73	0.44	0.79	0.41	-1.23	-0.12
s3e6	0	0.73	0.44	0.69	0.46	0.93	0.10	s6e14	3	0.38	0.49	0.42	0.50	-0.75	-0.07
s3e8	0	0.87	0.34	0.96	0.19	-2.59*	-0.35	s9e1	2	0.36	0.48	0.35	0.48	0.32	0.04
s3e9	0	0.73	0.45	0.77	0.42	-0.87	-0.10	s9e2	2	0.26	0.44	0.32	0.47	-1.15	-0.14
s3e11	1	0.13	0.34	0.17	0.38	-0.94	-0.10	s9e4	0	0.57	0.50	0.54	0.50	0.58	0.05
s3e12	3	0.13	0.34	0.04	0.19	3.48*	0.35	s9e5	1	0.22	0.42	0.32	0.47	-1.83	-0.22
s3e13	1	0.30	0.46	0.27	0.45	0.47	0.06	s9e6	0	0.43	0.50	0.39	0.49	0.96	0.09
s3e14	1	0.26	0.44	0.32	0.47	-1.15	-0.14	s9e7	2	0.39	0.49	0.35	0.48	0.73	0.09
s3e15	0	0.68	0.47	0.50	0.50	3.21*	0.36	s9e8	2	0.33	0.47	0.34	0.47	-0.16	-0.02
s4e1	2	0.37	0.49	0.34	0.47	0.60	0.07	s9e9	3	0.64	0.48	0.72	0.45	-1.68	-0.17
s4e2	3	0.68	0.47	0.72	0.45	-0.71	-0.08	s9e11	0	0.60	0.49	0.60	0.49	0.00	0.00
s4e3	0	0.78	0.42	0.81	0.40	-0.77	-0.07	s9e14	2	0.20	0.40	0.20	0.40	0.00	0.00
s4e4	0	0.40	0.49	0.39	0.49	0.21	0.02	s9e15	2	0.31	0.46	0.24	0.43	1.42	0.16
s4e5	0	0.07	0.26	0.10	0.30	-0.90	-0.10								

Anmerkungen:

N= 113. AdO= Angabe der Originalperson.

M= Mittelwert; S= Standardabweichung. t= T-Test bei verbundenen Stichproben, df=112. d= Effektstärke.

*Auf dem Niveau 0.05 signifikanter Unterschied zwischen den Mittelwerten „Text“ und „Video“.

Tabelle A-35: Mittelwertvergleich der text- und der videobasierten Forschungsversion (ViTEIP; Scoring: Konsens-Anteil)

Item	AdO	Konsens-Anteil				Mittelwert- vergleich Text und Video		Item	AdO	Konsens-Anteil				Mittelwert- vergleich Text und Video	
		Text		Video						Text		Video			
		M	S	M	S	t	d			M	S	M	S	t	d
s1e1	3	0.31	0.08	0.32	0.12	-0.25	-0.03	s4e6	3	0.66	0.26	0.67	0.28	-0.62	-0.06
s1e2	3	0.35	0.12	0.36	0.11	-0.76	-0.09	s4e7	3	0.81	0.25	0.82	0.24	-0.51	-0.06
s1e3	0	0.69	0.28	0.54	0.23	5.48*	0.56	s4e8	1	0.35	0.12	0.32	0.11	2.17*	0.22
s1e4	0	0.37	0.15	0.35	0.16	1.50	0.14	s4e9	0	0.38	0.14	0.34	0.13	3.42*	0.28
s1e5	0	0.29	0.09	0.30	0.09	-1.16	-0.13	s4e11	0	0.67	0.29	0.61	0.27	2.15*	0.22
s1e6	2	0.34	0.13	0.50	0.21	-7.97*	-0.91	s4e14	0	0.26	0.06	0.27	0.07	-0.66	-0.08
s1e7	1	0.28	0.09	0.27	0.07	0.91	0.11	s6e1	0	0.30	0.09	0.37	0.16	-4.44*	-0.49
s1e8	2	0.39	0.15	0.41	0.16	-1.20	-0.12	s6e2	0	0.30	0.10	0.30	0.08	0.20	0.02
s1e9	0	0.28	0.09	0.30	0.08	-1.47	-0.18	s6e3	0	0.40	0.16	0.37	0.14	2.18*	0.21
s1e12	0	0.74	0.26	0.73	0.26	0.51	0.05	s6e4	0	0.47	0.22	0.45	0.14	1.13	0.12
s1e13	0	0.88	0.22	0.77	0.26	4.46*	0.48	s6e5	0	0.29	0.08	0.33	0.07	-3.82*	-0.43
s3e1	0	0.34	0.12	0.30	0.08	4.20*	0.44	s6e6	3	0.31	0.09	0.31	0.08	0.27	0.03
s3e2	0	0.30	0.08	0.29	0.07	1.56	0.20	s6e7	2	0.55	0.21	0.44	0.19	5.74*	0.56
s3e4	0	0.29	0.06	0.27	0.07	1.62	0.20	s6e9	0	0.95	0.15	0.89	0.21	3.10*	0.32
s3e5	0	0.29	0.08	0.34	0.08	-6.70*	-0.70	s6e12	0	0.58	0.25	0.65	0.27	-2.82*	-0.27
s3e6	0	0.59	0.24	0.54	0.23	1.92	0.20	s6e14	3	0.35	0.12	0.31	0.12	3.36*	0.35
s3e8	0	0.77	0.28	0.93	0.17	-4.91*	-0.67	s9e1	2	0.28	0.08	0.27	0.07	0.42	0.04
s3e9	0	0.58	0.25	0.64	0.26	-2.22*	-0.25	s9e2	2	0.33	0.13	0.32	0.11	0.76	0.07
s3e11	1	0.64	0.29	0.51	0.25	4.82*	0.48	s9e4	0	0.42	0.18	0.40	0.16	1.38	0.12
s3e12	3	0.29	0.09	0.32	0.08	-4.72*	-0.43	s9e5	1	0.29	0.09	0.28	0.08	1.49	0.18
s3e13	1	0.40	0.17	0.42	0.18	-1.89	-0.15	s9e6	0	0.32	0.11	0.32	0.10	0.75	0.06
s3e14	1	0.31	0.13	0.37	0.15	-4.05*	-0.39	s9e7	2	0.31	0.09	0.29	0.07	2.07*	0.23
s3e15	0	0.51	0.26	0.37	0.16	6.02*	0.66	s9e8	2	0.27	0.07	0.28	0.07	-0.47	-0.05
s4e1	2	0.29	0.10	0.28	0.08	0.86	0.10	s9e9	3	0.49	0.20	0.56	0.26	-2.88*	-0.28
s4e2	3	0.52	0.23	0.57	0.24	-1.94	-0.20	s9e11	0	0.41	0.23	0.42	0.22	-0.43	-0.03
s4e3	0	0.65	0.27	0.68	0.27	-1.44	-0.13	s9e14	2	0.31	0.12	0.37	0.15	-4.47*	-0.44
s4e4	0	0.31	0.11	0.30	0.09	0.25	0.02	s9e15	2	0.32	0.12	0.35	0.18	-2.22*	-0.19
s4e5	0	0.34	0.14	0.33	0.13	0.55	0.06								

Anmerkungen:

N= 113. AdO= Angabe der Originalperson.

M= Mittelwert; S= Standardabweichung. t= T-Test bei verbundenen Stichproben, df=112. d= Effektstärke.

*Auf dem Niveau 0.05 signifikanter Unterschied zwischen den Mittelwerten „Text“ und „Video“.

Tabelle A-36: Mittelwertvergleich der text- und der videobasierten Forschungsversion (ViTEIP; Scoring: Konsens-Modal)

Item	AdO	Konsens-Modal				Mittelwert- vergleich Text und Video		Item	AdO	Konsens-Modal				Mittelwert- vergleich Text und Video	
		Text		Video						Text		Video			
		M	S	M	S	t	d			M	S	M	S	t	d
s1e1	3	0.36	0.48	0.42	0.50	-0.93	-0.13	s4e6	3	0.79	0.41	0.81	0.40	-0.42	-0.04
s1e2	3	0.43	0.50	0.47	0.55	-0.46	-0.07	s4e7	3	0.89	0.31	0.90	0.30	-0.24	-0.03
s1e3	0	0.81	0.39	0.70	0.46	2.56*	0.27	s4e8	1	0.41	0.49	0.42	0.50	-0.26	-0.04
s1e4	0	0.50	0.50	0.50	0.50	0.00	0.00	s4e9	0	0.50	0.50	0.46	0.50	1.00	0.09
s1e5	0	0.36	0.48	0.36	0.48	0.00	0.00	s4e11	0	0.81	0.39	0.76	0.43	1.23	0.13
s1e6	2	0.44	0.50	0.65	0.48	-3.83*	-0.43	s4e14	0	0.31	0.46	0.35	0.48	-1.00	-0.09
s1e7	1	0.36	0.48	0.34	0.47	0.40	0.06	s6e1	0	0.39	0.49	0.52	0.50	-2.60*	-0.27
s1e8	2	0.52	0.50	0.55	0.50	-0.52	-0.05	s6e2	0	0.41	0.49	0.36	0.48	0.69	0.09
s1e9	0	0.39	0.49	0.36	0.48	0.42	0.05	s6e3	0	0.53	0.50	0.50	0.50	0.73	0.07
s1e12	0	0.85	0.36	0.84	0.37	0.24	0.02	s6e4	0	0.63	0.49	0.57	0.50	1.26	0.13
s1e13	0	0.94	0.24	0.87	0.34	2.17*	0.24	s6e5	0	0.32	0.47	0.39	0.49	-1.27	-0.15
s3e1	0	0.44	0.50	0.36	0.48	1.22	0.16	s6e6	3	0.39	0.49	0.37	0.49	0.32	0.04
s3e2	0	0.37	0.49	0.64	0.48	-4.31*	-0.55	s6e7	2	0.69	0.46	0.59	0.49	2.00*	0.20
s3e4	0	0.32	0.47	0.33	0.47	-0.15	-0.02	s6e9	0	0.97	0.16	0.94	0.24	1.65	0.17
s3e5	0	0.35	0.48	0.42	0.50	-1.41	-0.16	s6e12	0	0.73	0.44	0.80	0.40	-1.47	-0.15
s3e6	0	0.73	0.44	0.69	0.46	0.93	0.10	s6e14	3	0.43	0.50	0.42	0.50	0.24	0.04
s3e8	0	0.87	0.34	0.96	0.19	-2.59*	-0.35	s9e1	2	0.36	0.48	0.35	0.48	0.32	0.04
s3e9	0	0.73	0.45	0.79	0.41	-1.26	-0.14	s9e2	2	0.46	0.50	0.42	0.50	0.96	0.09
s3e11	1	0.78	0.42	0.68	0.47	2.24*	0.22	s9e4	0	0.57	0.50	0.55	0.50	0.39	0.04
s3e12	3	0.36	0.48	0.37	0.49	-0.13	-0.02	s9e5	1	0.39	0.49	0.34	0.47	0.88	0.11
s3e13	1	0.54	0.50	0.58	0.50	-0.82	-0.07	s9e6	0	0.43	0.50	0.41	0.49	0.58	0.05
s3e14	1	0.44	0.50	0.50	0.50	-1.03	-0.11	s9e7	2	0.39	0.49	0.35	0.48	0.73	0.09
s3e15	0	0.68	0.47	0.51	0.50	3.08*	0.35	s9e8	2	0.33	0.47	0.34	0.47	-0.16	-0.02
s4e1	2	0.37	0.49	0.34	0.47	0.60	0.07	s9e9	3	0.64	0.48	0.72	0.45	-1.68	-0.17
s4e2	3	0.68	0.47	0.72	0.45	-0.71	-0.08	s9e11	0	0.60	0.49	0.60	0.49	0.00	0.00
s4e3	0	0.78	0.42	0.81	0.40	-0.77	-0.07	s9e14	2	0.43	0.50	0.51	0.50	-1.53	-0.16
s4e4	0	0.40	0.49	0.39	0.49	0.21	0.02	s9e15	2	0.43	0.50	1.20	1.62	-4.66*	-0.64
s4e5	0	0.45	0.50	0.44	0.50	0.15	0.02								

Anmerkungen:

N= 113. AdO= Angabe der Originalperson.

M= Mittelwert; S= Standardabweichung. t= T-Test bei verbundenen Stichproben, df=112. d= Effektstärke.

*Auf dem Niveau 0.05 signifikanter Unterschied zwischen den Mittelwerten „Text“ und „Video“.

Tabelle A-37: Deskriptive Statistiken von „Text“ und „Video“ bei unterschiedlichen Scorings (ViTEIP)

Präsentations- modus	M				s ²				α			
	T-Ü	T-T	K-A	K-M	T-Ü	T-T	K-A	K-M	T-Ü	T-T	K-A	K-M
Text	119.64	24.62	23.72	29.51	96.29	26.79	5.64	31.56	0.68	0.63	0.73	0.64
Video	117.37	23.70	23.60	30.51	105.77	32.30	4.57	25.13	0.71	0.70	0.67	0.44

Anmerkungen:

Scoring: T-Ü= Target-Übereinstimmung; T-T= Target-Treffer; K-A= Konsens-Anteil; K-M= Konsens-Modal. Für eine detaillierte Beschreibung der Scorings siehe Kapitel 3.3.1

M= Mittelwert. s²= Varianz. α= interne Konsistenz Cronbachs α, berechnet über Items.

Tabelle A-38: Vergleich der Akzeptanzwerte der text- und der videobasierten Forschungsversion (ViTEIP)

Kriterien ¹	„Text“		„Video“		Mittelwertvergleich: „Text“ und „Video“	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>d</i>
Nützlichkeit	1.72	0.81	1.70	0.75	0.17	0.03
Spas	1.76	0.80	1.81	0.83	-0.39	-0.06
Akzeptanz	1.58	0.82	1.56	0.88	0.23	0.02
Fairness	1.42	0.84	1.35	0.82	0.61	0.08

Anmerkungen:

M= Mittelwert; *SD*= Standardabweichung. *t*= T-Test bei verbundenen Stichproben, *df*=112. *d*= Effektstärke.

*Auf dem Niveau 0.05 signifikant.

¹ Beantwortung auf vierstufiger Antwortskala mit „0=stimme nicht zu“, „1=stimme eher nicht zu“, „2=stimme eher zu“, „3=stimme zu“.

Tabelle A-39: Korrelationsmatrix ViTEIP „Text“ und „Video“ bei unterschiedlichen Scorings

	„Video“			
	Target-Übereinstimmung	Target-Treffer	Konsens-Anteil	Konsens-Modal
„Text“ Target-Übereinstimmung	0.69**	0.63**	0.44**	0.40**
Target-Treffer	0.68**	0.70**	0.51**	0.44**
Konsens-Anteil	0.51**	0.48**	0.71**	0.58**
Konsens-Modal	0.53**	0.49**	0.64**	0.58**

Anmerkungen:

Korrelation nach Pearson, einseitige Testung, **p*<0.05, ***p*<0.01.

Für eine detaillierte Beschreibung der Scorings siehe Kapitel 3.3.1

Anhang K. Die Geschichte mit dem Hammer (Watzlawick, 2004)

Ein Mann will ein Bild aufhängen. Den Nagel hat er, nicht aber den Hammer. Der Nachbar hat einen. Also beschließt unser Mann, hinüberzugehen und ihn auszuborgen. Doch da kommt ihm ein Zweifel: Was, wenn der Nachbar mir den Hammer nicht leihen will? Gestern schon grüßte er mich nur so flüchtig. Vielleicht war er in Eile. Aber vielleicht war die Eile nur vorgeschützt, und er hat etwas gegen mich. Und was? Ich habe ihm nichts angetan; der bildet sich da etwas ein. Wenn jemand von mir ein Werkzeug borgen wollte, *ich* gäbe es ihm sofort. Und warum er nicht? Wie kann man einem Mitmenschen einen so einfachen Gefallen abschlagen? Leute wie dieser Kerl vergiften einem das Leben. Und dann bildet er sich noch ein, ich sei auf ihn angewiesen. Bloß weil er einen Hammer hat. Jetzt reicht's mir wirklich. – Und so stürmt er hinüber, läutet, der Nachbar öffnet, doch bevor er "Guten Tag" sagen kann, schreit ihn unser Mann an: "Behalten Sie sich Ihren Hammer, Sie Rüpel!"

(Watzlawick, 2004)

7 Literaturverzeichnis

- Abdi, H. (2007). Signal Detection Theory. In N. Salkind (Ed.), *Encyclopedia of Measurement and Statistics* (pp. 886-889). Thousand Oaks (CA): Sage.
- Allen, V. D., Weissmann, A., Hellwig, S., MacCann, C., & Roberts, R. D. (2014). Development of the situational test of emotional understanding – brief (STEU-B) using item response theory. *Personality and Individual Differences*, 65, pp. 3-7.
- Amelang, M., & Steinmayr, R. (2006). Is there a validity increment for tests of emotional intelligence in explaining the variance of performance criteria? *Intelligence*, 34, pp. 459-468.
- Amthauer, R., Brocke, B., Liepmann, D., & Beauducel, A. (2001). *Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R)*. Göttingen: Hogrefe.
- Antonakis, J., Ashkanasy, N. M., & Dasbourough, M. T. (2009). Does leadership need emotional intelligence? *The Leadership Quarterly*, 20, pp. 247-261.
- Austin, E. J. (2004). An investigation of the relationship between trait emotional intelligence and emotional task performance. *Personality and Individual Differences*, 36, pp. 1855–1864.
- Austin, E. J. (2010). Measurement of ability emotional intelligence: Results for two new tests. *British Journal of Psychology*, 101, pp. 563-578.
- Austin, E. J., & Saklofske, D. H. (2006). Viel zu viele Intelligenzen? Über die Gemeinsamkeiten und Unterschiede zwischen sozialer, praktischer und emotionaler Intelligenz. In R. Schulze, P. A. Freund, & R. D. Roberts, *Emotionale Intelligenz. Ein Internationales Handbuch* (S. 117-137). Göttingen: Hogrefe.
- Austin, E. J., Farrelly, D., Black, C., & Moore, H. (2007). Emotional intelligence, Machiavellianism and emotional manipulation: Does EI have a dark side? *Personality and Individual Differences*, 43, pp. 179-189.
- Bäckström, M., Björklund, F., & Larsson, M. R. (2009). Five-factor inventories have a major general factor related to social desirability which can be reduced by framing items neutrally. *Journal of Research in Personality*, 43, pp. 335–344.
- Bagby, R. M., Parker, J. D., & Taylor, G. J. (1994). The Twenty-Item Toronto Alexithymia Scale-1: Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research*, 38, pp. 23-32.
- Bänziger, T., Grandjean, D., & Scherer, K. R. (2009). Emotion Recognition From Expressions in Face, Voice, and Body: The Multimodal Emotion Recognition Test (MERT). *Emotion*, 9 (5), pp. 691–704.
- Barchard, K. A. (2003). Does Emotional Intelligence Assist in the Prediction of Academic Success? *Educational and Psychological Measurement*, 63 (5), pp. 840-858.

- Barchard, K. A., & Russell, J. A. (2006). Bias in consensus scoring, with examples from ability emotional intelligence tests. *Psicothema*, 18, pp. 49-54.
- Barchard, K. A., Hensley, S., & Anderson, E. (2013). When proportion consensus scoring works. *Personality and Individual Differences*, 55, pp. 14-18.
- Bar-On, R. (1997). *Bar-On Emotional Quotient Inventory: Technical manual*. Toronto, Canada: Mult-Health Systems.
- Bar-On, R. (2000). Emotional and social intelligence: Insights from the Emotional Quotient Inventory (EQ-i). In R. Bar-On, & J. D. Parker (Eds.), *Handbook of emotional intelligence: Theory, development, assessment and application at home, school and in the workplace*. (pp. 363-388). San Francisco: Jossey-Bass.
- Bar-On, R., & Handley, R. (2003). *The Bar-On EQ-360: Technical manual*. Toronto, Canada: Multi-Health Systems.
- Batista-Foguet, J. M., Saris, W., Boyatzis, R., Guillén, L., & Serlavós, R. (2009). Effect of response scale on assesement of emotional intelligence competencies. *Personality and Individual Differences*, 46, pp. 575-580.
- Beaupré, M. G., Cheung, N., & Hess, U. (2000). *The Montreal Set of Facial Displays of Emotion*. Montreal: University of Quebec at Montreal.
- Beauvais, A. M., Brady, N., O'Shea, E. R., & Griffin, M. T. (2011). Emotional intelligence and nursing performance among nursing students. *Nurse Education Today*, 31, pp. 396-401.
- Bergman, M. E., Drasgow, F., Donovan, M. A., Henning, J. B., & Juraska, S. E. (2006). Scoring Situational Judgment Tests: Once You Get the Data, Your Troubles Beginn. *International Journal of Selection and Assessment*, 14 (3), pp. 223-235.
- Billiet, J. B., & McClendon, M. J. (2000). Modeling Acquiescence in Measurement Models for Two Balanced Sets of Items. *Structural Equation Modeling*, 7, pp. 608-628.
- Blickle, G., Kramer, J., & Mierke, J. (2010). Telephone-Administered Intelligence Testing for Research in Work and Organizational Psychology. A Comparative Assessment Study. *European Journal of Psychological Assessment*, 26 (3), pp. 154-161.
- Blickle, G., Momm, T. S., Kramer, J., Mierke, J., Liu, Y., & Ferris, G. F. (2009). Construct and Criterion-Related Validation of a Measure of Emotional Reasoning Skills: A two-study investigation. *International Journal of Selection and Assesement*, 17 (1), pp. 101-118.
- Blickle, G., Momm, T., Liu, Y., Witzki, A., & Steinmayr, R. (2011). Construct Validation of the Test of Emotional Intelligence (TEMINT). A Two-Study Investigation. *European Journal of Psychological Assessment*, 27 (4), pp. 282-289.
- Borkenau, P., & Ostendorf, F. (1993). *NEO-Fünf-Faktoren Inventar nach Costa und McCrae (NEO-FFI)*. Göttingen: Hogrefe.

- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler (6. Auflage)*. Heidelberg: Springer Medizin Verlag.
- Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Heidelberg: Springer.
- Bowman, D. B., Markham, P. M., & Roberts, R. D. (2001). Expanding the frontier of human cognitive abilities: so much more than (plain) g! *Learning and Individual Differences*, 13 (2), pp. 127–158.
- Boyatzis, R. E., Goleman, D., & Rhee, K. (2000). Clustering competence in emotional intelligence: Insights from the Emotional Competence Inventory (ECI)s. In R. Bar-On, & J. A. Parker (Eds.), *Handbook of emotional intelligence* (pp. 343-362). San Francisco: Jossey-Bass.
- Brackett, M. A., & Mayer, J. D. (2003). Convergent, Discriminant, and Incremental Validity of Competing Measures of Emotional Intelligence. *Personality and Social Psychology Bulletin*, 29, pp. 1147-1158.
- Brackett, M. A., Mayer, J. D., & Warner, R. M. (2004). Emotional intelligence and its relation to everyday behaviour. *Personality and Individual Differences*, 36, pp. 1387–1402.
- Brackett, M. A., Rivers, S. E., Lerner, N., Salovey, P., & Shiffman, S. (2006). Relating Emotional Abilities to Social Functioning: A Comparison of Self-Report and Performance Measures of Emotional Intelligence. *Journal of Personality and Social Psychology*, 91 (4), pp. 780–795.
- Brickenkamp, R. (2002). *Aufmerksamkeits-Belastungstest d2 (9., überarbeitete und neu normierte Auflage.)*. Göttingen: Hogrefe.
- Brickenkamp, R., Schmidt-Atzert, L., & Liepmann, D. (2010). *Test d2-R: Aufmerksamkeits-Konzentrations-Test, revidierte Form*. Göttingen: Hogrefe.
- Brody, N. (2004). What cognitive intelligence is and what emotional intelligence is not. *Psychological Inquiry*, 15, pp. 234-238.
- Brody, N. (2009). Beyond g. In K. R. Murphy (Ed.), *A Critique of Emotional Intelligence. What are the problems and how can they be fixed?* (pp. 161-185). New York: Routledge.
- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion, 2., aktualisierte Auflage*. München: Pearson Studium.
- Carroll, J. M., & Russell, J. A. (1996). Do Facial Expressions Signal Specific Emotions? Judging Emotion From the Face in Context. *Journal of Personality and Social Psychology*, 70 (2), pp. 205-218.
- Cattell, R. B. (1978). *Die empirische Erforschung der Persönlichkeit*. Weinheim: Beltz.
- Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology*, 82, pp. 143-159.

- Choi, S., & Kluemper, D. H. (2012). The relative utility of differing measures of emotional intelligence: Other-report EI as a predictor of social functioning. *Revue européenne de psychologie appliquée*, 62, pp. 121–127.
- Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology*, 63, pp. 83-117.
- Christiansen, N. D., Janovics, J. E., & Siers, B. P. (2010). Emotional Intelligence in Selection Contexts: Measurement method, criterion-related validity, and vulnerability to response distortion. *International Journal of Selection and Assessment*, 18(1), pp. 87-101.
- Ciarrochi, J., Caputi, P., & Mayer, J. D. (2003). The distinctiveness and utility of a measure of trait emotional awareness. *Personality and Individual Differences*, 34, pp. 1477-1490.
- Ciarrochi, J., Chan, A., & Caputi, P. (2000). A critical evaluation of the emotional intelligence construct. *Personality and Individual Differences*, 28, pp. 539-561.
- Ciarrochi, J., Chan, A., Caputi, P., & Roberts, R. (2001). Measuring Emotional Intelligence. In J. Ciarrochi, J. P. Forgas, & J. D. Mayer (Eds.), *Emotional Intelligence in Everyday Life* (pp. 25-45). New York: Psychology Press.
- Ciarrochi, J., Forgas, J. P., & Mayer, J. D. (2001). *Emotional Intelligence In Everyday Life*. New York: Psychology Press.
- Cohen, J. (1992). A Power Primer. *Psychological Bulletin*, 112 (1), pp. 155-159.
- Conzelmann, K., Weis, S., & Süß, H.-M. (2013). New Findings About Social Intelligence: Development and Application of the Magdeburg Test of Social Intelligence (MTSI). *Journal of Individual Differences*, 34 (3), pp. 119–137.
- Cortina, J. M. (1993). What is Coefficient Alpha? An Examination of Theory and Applications. *Journal of Applied Psychology*, 78 (1), pp. 98-104.
- Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five Factor Inventory. Professional Manual*. Odessa, Florida: Psychological Assessment Resources.
- Costanzo, M., & Archer, D. (1993). *Interpersonal Perception Task-15*. Berkeley, CA: University of California Extension Media Center.
- Cote, J. A., & Buckley, R. (1987). Estimating Trait, Method, and Error Variance: Generalizing Across 70 Construct Validation Studies. *Journal of Marketing Research*, 3, pp. 315-318.
- Davies, M., Stankov, L., & Roberts, R. D. (1998). Emotional intelligence: In search of an elusive construct. *Journal of Personality and Social Psychology*, 75, pp. 989-1015.
- Davis, S. K., & Humphrey, N. (2012). Emotional intelligence predicts adolescent mental health beyond personality and cognitive ability. *Personality and Individual Differences*, 52, pp. 144–149.

- Day, A. L., & Carroll, S. A. (2004). Using an ability-based measure of emotional intelligence to predict individual performance, group performance, and group citizenship behaviours. *Personality and Individual Differences*, 36, pp. 1443–1458.
- de Meijer, L. A., Born, M. P., van Zielst, J., & van der Molen, H. T. (2010). Construct-Driven Development of a Video-Based Situational Judgment Test for Integrity. *European Psychologist*, 15 (3), pp. 229-236.
- DeBusk, K. P., & Austin, E. J. (2011). Emotional intelligence and social perception. *Personality and Individual Differences*, 51, pp. 764-768.
- Ekman, P., & Friesen, W. V. (1967). Head and Body Cues in the Judgment of Emotion: a Reformulation. *Perceptual and Motor Skills*, 24, pp. 711-724.
- Fallon, C. K., Panganiban, A. R., Wohleber, R., Matthews, G., Kustubayeva, A. M., & Roberts, R. (2014). Emotional intelligence, cognitive ability and information search in tactical decision-making. *Personality and Individual Differences*, 65, pp. 24–29.
- Fan, H., Jackson, T., Yang, X., Tang, W., & Zhang, J. (2010). The factor structur of the Mayer-Salovey-Caruso Emotional Intelligence Test V2.0 (MSCEIT): a meta-analytic structural equation modeling approach. *Personality and Individual Differences*, 48, pp. 781-785.
- Farrelly, D., & Austin, E. J. (2011). Ability EI as an intelligence? Associations of the MSCEIT with performance on emotion processing and social tasks and with cognitive ability. *Cognition & Emotion*, 21 (5), pp. 1043-1063.
- Ferguson, F. J., & Austin, E. J. (2011). The factor structures of the STEM and the STEU. *Personality and Individual Differences*, 51, pp. 791-794.
- Fiori, M., & Antonakis, J. (2011). The ability model of emotional intelligence: Searching for valid measures. *Personality and Individual Differences*, 50, pp. 329-334.
- Fiori, M., & Antonakis, J. (2012). Selective attention to emotional stimuli: What IQ and openness do, and emotional intelligence does not. *Intelligence*, 40, pp. 245-254.
- Flanagan, D. P., & Dixon, S. G. (2014, Januar 22). *Encyclopedia of Special Education: The Cattell-Horn-Carroll Theory of Cognitive Abilities*. Retrieved from Wiley Online Library: DOI: 10.1002/9781118660584.es0431
- Flanagan, J. C. (1954). The Critical Incident Technique. *Psychological Bulletin*, 51 (4), pp. 327-358.
- Follesdal, H., & Hagtvet, K. A. (2009). Emotional Intelligence: The MSCEIT from the perspective of generalizability theory. *Intelligence*, 37, pp. 94-105.
- Forgas, J. P. (2007). When sad is better than happy: Negative affect can improve the quality and effectiveness of persuasive messages and social influence strategies. *Journal of Experimental Social Psychology*, 43, pp. 513–528.
- Forgas, J. P., & East, R. (2008). On being happy and gullible: Mood effects on skepticism and the detection of deception. *Journal of Experimental Social Psychology*, 44, pp. 1362–1367.

- Forgas, J. P., Laham, S. M., & Vargas, P. T. (2005). Mood effects on eyewitness memory: Affective influences on susceptibility to misinformation. *Journal of Experimental Social Psychology*, 41, pp. 574–588.
- Fridlund, A. J. (1991). Sociality of solitary smiling: Potentiation by an implicit audience. *Journal of Personality and Social Psychology*, 60 (2), pp. 229-240.
- Furnham, A. (2001). Self-estimates of intelligence: Culture and gender differences in self and other estimates of both general (g) and multiple intelligences. *Personality and Individual Differences*, 31, pp. 1381-1405.
- Furnham, A., McClelland, A., & Mansi, A. (2012). Selecting your boss: sex, age, IQ and EQ factors. *Personality and Individual Differences*, 53, pp. 552-556.
- Gardner, K. J., & Qualter, P. (2011). Factor structure, measurement invariance and structural invariance of the MSCEIT V2.0. *Personality and Individual Differences*, 51, pp. 492-496.
- Geher, G., Warner, R. M., & Brown, A. S. (2001). Predictive validity of the emotional accuracy research scale. *Intelligence*, 29, pp. 373-388.
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & et al. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40, pp. 84-96.
- Goleman, D. (1995). *Emotional intelligence*. New York: Bantam.
- Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley.
- Greenleaf, E. A. (1992). Improving rating scale measures by detecting and correcting bias components in some response styles. *Journal of Marketing Research*, 29, pp. 176-188.
- Guttman, L., & Levy, S. (1991). Two structural laws for intelligence. *Intelligence*, 15, pp. 79-103.
- Holtus, P. (2008). *Entwicklung eines videobasierten Testverfahrens zur Erfassung der Emotionalen Intelligenz von Mitarbeitern im Pflegeberuf*. Unveröffentlichte Diplomarbeit. Marburg: Philipps-Universität.
- Holz-Ebeling, F., & Steinmetz, M. (1994). Entwicklung eines situationsbezogenen Empathiefragebogens mittels experimenteller Konstruktionsprinzipien. *Zeitschrift für Sozialpsychologie*, 25, S. 155-169.
- Horn, J. L., & Blankson, N. A. (2012). Foundations for Better Understanding of Cognitive Abilities. In D. P. Flanagan, & P. L. Harrison (Eds.), *Contemporary Intellectual Assessment. Third Edition*. (pp. 73-98). New York: The Guilford Press.
- Jacobs, K. E., Szer, D., & Roodenburg, J. (2012). The moderating effect of personality on the accuracy of self-estimates of intelligence. *Personality and Individual Differences*, 52, pp. 744–749.
- Joseph, D. L., & Newman, D. A. (2010). Emotional Intelligence: An Integrative Meta-Analysis and Cascading Model. *Journal of Applied Psychology*, 95 (1), pp. 54–78.

- Kang, S.-M., Day, J. D., & Meara, N. M. (2006). Soziale und emotionale Intelligenz: Gemeinsamkeiten und Unterschiede. In R. Schulz, P. A. Freund, & R. D. Roberts (Hrsg.), *Emotionale Intelligenz: Ein Internationales Handbuch* (S. 101-115). Göttingen: Hogrefe.
- Kanning, U. P., Grewe, K., Hollenberg, S., & Hadouch, M. (2006). From the Subjects' Point of View: Reactions to Different Types of Situational Judgment Items. *European Journal of Psychological Assessment*, 22 (3), pp. 168-176.
- Keele, S. M., & Bell, R. C. (2009). Consensus scoring, correct responses and reliability of the MSCEIT V2. *Personality and Individual Differences*, 47, pp. 740-747.
- Kersting, M., Althoff, K., & Jäger, A. (2008). *WIT-2. Der Wilde-Intelligenztest*. Göttingen: Hogrefe.
- Landy, F. J. (2006). The Long, Frustrating, and Fruitless Search for Social Intelligence: A Cautionary Tale. In K. R. Murphy (Ed.), *A critique of emotional intelligence: What are the problems and how can they be fixed?* (pp. 81-123). Mahwah: Lawrence Erlbaum Associates.
- Lane, R., Quinlan, D., Schwartz, G. E., Walker, P., & Zeitlin, S. (1990). The levels of emotional awareness scale: A cognitive-developmental measure of emotion. *Journal of Personality Assessment*, 55, pp. 124-134.
- Legree, P. J. (1995). Evidence for an Oblique Social Intelligence Factor Established With a Likert-Based Testing Procedure. *Intelligence*, 21, pp. 247-266.
- Legree, P. J., Psotka, J., Tremble, T. R., & Bourne, D. (2005). Using Consensus Based Measurement to Assess Emotional Intelligence. In R. Schulze, & R. Roberts (Eds.), *International Handbook of Emotional Intelligence* (pp. 99-123). Berlin, Germany: Hogrefe & Huber.
- Libbrecht, N., & Lievens, F. (2012). Validity evidence for the situational judgment test paradigm in emotional intelligence measurement. *International Journal of Psychology*, 47 (6), pp. 438-447.
- Libbrecht, N., Lievens, F., Carette, B., & Côté, S. (2014). Emotional Intelligence Predicts Success in Medical School. *Emotion*, 14 (1), pp. 64-73.
- Lievens, F. (2013). Adjusting medical school admission: assessing interpersonal skills using situational judgment tests. *Medical Education*, 47, pp. 182-189.
- Lievens, F., & Sackett, P. R. (2006). Video-Based Versus Written Situational Judgment Tests: A Comparison in Terms of Predictive Validity. *Journal of Applied Psychology*, 91, pp. 1181-1188.
- Lindebaum, D. (2013). Does emotional intelligence moderate the relationship between mental health and job performance? An exploratory study. *European Management Journal*, 31, pp. 538-548.
- Lishner, D. A., Swim, E. R., Hong, P. Y., & Vitacco, M. J. (2011). Psychopathy and ability emotional intelligence: Widespread or limited association among facets? *Personality and Individual Differences*, 50, pp. 1029-1033.

- Livingstone, H. A., & Day, A. L. (2005). Comparing the construct and criterion-related validity of ability-based and mixed-model measures of emotional intelligence. *Educational and Psychological Measurement*, 65 (5), pp. 757-779.
- Locke, E. A. (2005). Why emotional intelligence is an invalid concept. *Journal of Organizational Behavior*, 26, pp. 425-431.
- Lopes, P. N., Brackett, M. A., Nezlek, J. B., Schütz, A., Sellin, I., & Salovey, P. (2004). Emotional Intelligence and Social Interaction. *Personality and Social Psychology Bulletin*, 30 (8), pp. 1018-1034.
- Lopes, P. N., Salovey, P., & Straus, R. (2003). Emotional intelligence, personality, and the perceived quality of social relationships. *Personality and Individual Differences*, 35, pp. 641-658.
- MacCann, C. E., & Roberts, R. D. (2008). New Paradigms for Assessing Emotional Intelligence: Theory and Data. *Emotion*, 8 (4), pp. 540-551.
- MacCann, C. E., Roberts, R. D., Matthews, G., & Zeidner, M. (2004). Consensus scoring and empirical option weighting of performance-based emotional intelligence. *Personality and Individual Differences*, 36, pp. 645-662.
- MacCann, C., Fogarty, G. J., Zeidner, M., & Roberts, R. D. (2011). Coping mediates the relationship between emotional intelligence (EI) and academic achievement. *Contemporary Educational Psychology*, 36, pp. 60-70.
- MacCann, C., Joseph, D. L., Newman, D. A., & Roberts, R. D. (2014). Emotional Intelligence Is a Second-Stratum Factor of Intelligence: Evidence From Hierarchical and Bifactor Models. *Emotion*, 14 (2), pp. 358-374.
- MacCann, C., Wang, L., Matthews, G., & Roberts, R. D. (2010). Emotional intelligence and the eye of the beholder: Comparing self- and parent-rated situational judgments in adolescents. *Journal of Research in Personality*, 44, pp. 673-676.
- Marentette, B. J., Meyers, L. S., Hurtz, G. M., & Kuang, C. C. (2012). Order Effects on Situational Judgment Test Items: A case of construct-irrelevant difficulty. *International Journal of Selection and Assessment*, 20 (3), pp. 310-332.
- Martins, A., Ramalho, N., & Morin, E. (2010). A comprehensive meta-analysis of the relationship between Emotional Intelligence and health. *Personality and Individual Differences*, 49, pp. 554-564.
- Masuda, T., Ellsworth, P. C., Mesquita, B., Leu, J., Tanida, S., & Van de Veerdonk, E. (2008). Placing the Face in Context: Cultural Differences in the Perception of Facial Emotion. *Journal of Personality and Social Psychology*, 94 (3), pp. 365-381.
- Matthews, G., Emo, A. K., Roberts, R. D., & Zeidner, M. (2006). What Is This Thing Called Emotional Intelligence? In K. R. Murphey (Ed.), *A critique of emotional intelligence. What are the problems and how can they be fixed?* (pp. 3-36). New York: Routledge.

- Matthews, G., Roberts, R. D., & Zeidner, M. (2004). Seven myths about emotional intelligence. *Psychological Inquiry*, 15, pp. 179-196.
- Matthews, G., Zeidner, M., & Roberts, R. D. (2002). *Emotional Intelligence: Science and myth*. Cambridge, MA: MIT Press.
- Matthews, G., Zeidner, M., & Roberts, R. D. (2012). Emotional Intelligence: A promise unfulfilled? *Japanese Psychological Research*, 54 (2), pp. 105-127.
- Maul, A. (2011). The factor structure and cross-test convergence of the Mayer-Caruso model of emotional intelligence. *Personality and Individual Differences*, 50, pp. 457-463.
- Maul, A. (2012). The Validity of the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) as a Measure of Emotional Intelligence. *Emotion Review*, 4 (4), pp. 394-402.
- Mayer, J. D., & Geher, G. (1996). Emotional Intelligence and the Identification of Emotion. *Intelligence*, 22, pp. 89-133.
- Mayer, J. D., & Salovey, P. (1993). The intelligence of emotional intelligence. *Intelligence*, 17 (4), pp. 433-442.
- Mayer, J. D., & Salovey, P. (1997). What is emotional intelligence? In P. Salovey, & D. J. Sluyter, *Emotional development and emotional intelligence* (pp. 3-31). New York: Basic Books.
- Mayer, J. D., Caruso, D. R., & Salovey, P. (2000). Emotional intelligence meets traditional standards for an intelligence. *Intelligence*, 27 (4), pp. 267-298.
- Mayer, J. D., Salovey, P., & Caruso, D. R. (2002). *Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT): User's Manual*. Toronto: Mult-Health Systems.
- Mayer, J. D., Salovey, P., & Caruso, D. R. (2008). Emotional Intelligence: New Ability or Eclectic Traits? *American Psychologist*, 63 (6), pp. 503-517.
- Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2001). Emotional intelligence as a standard intelligence. *Emotion*, 1, pp. 232-242.
- Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2003). Measuring Emotional Intelligence With the MSCEIT V2.0. *Emotion*, 3 (1), pp. 97-105.
- McCrae, R. R. (2000). Emotional Intelligence from the perspective of the five-factor model of personality. In R. Bar-On, & J. D. Parker (Eds.), *The handbook of emotional intelligence* (pp. 263-276). San Francisco: Jossey-Bass.
- Mertin, I. (2004). *EQ-iBA. Validierung und psychometrische Evaluation eines Emotionalen Intelligenztests für den Berufsalltag*. Unveröffentlichte Diplomarbeit. Marburg: Philipps-Universität.
- Möller, J. H. (2010). *Kundenorientierung im Hotelfach. Die Entwicklung und Validierung eines Situational Judgment Tests. Wissenschaftliche Schriften der WWU Münster, Reihe VIII, Band 2*. Münster: Verlagshaus Monsenstein und Vannerdat.

- Moosbrugger, H. (2008). Klassische Testtheorie (KTT). In H. Moosbrugger, & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion* (pp. 99-112). Heidelberg: Springer.
- Moosbrugger, H., & Kelava, A. (2008). *Testtheorie und Fragebogenkonstruktion*. Berlin: Springer Verlag.
- Muck, P. M. (2013). Entwicklung von Situational Judgment Tests. *Zeitschrift für Arbeits- und Organisationspsychologie*, 57 (4), S. 185-205.
- Murphy, K. R., & Sideman, L. (2006). The Two Els. In K. R. Murphy (Ed.), *A Critique of Emotional Intelligence: What Are the Problems and How Can They Be Fixed?* (pp. 37-58). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Neubauer, A. C., & Freudenthaler, H. H. (2006). Modelle emotionaler Intelligenz. In R. Schulze, P. A. Freund, & R. D. Roberts, *Emotionale Intelligenz: Ein Internationales Handbuch* (S. 39-59). Göttingen: Hogrefe.
- Nowicki, S., & Duke, M. P. (1994). Individual Differences in the Nonverbal Communication of Affect: The diagnostic analysis of nonverbal accuracy scale. *Journal of Nonverbal Behavior*, 18, pp. 9-35.
- Palmer, B. R., Gignac, G., Manocha, R., & Stough, C. (2005). A psychometric evaluation of the Mayer-Salovey-Caruso Emotional Intelligence Test Version 2.0. *Intelligence*, 33, pp. 285-305.
- Parker, J. D. (2006). Die Relevanz emotionaler Intelligenz für die klinische Psychologie. In R. Schulze, P. A. Freund, & R. D. Roberts (Hrsg.), *Emotionale Intelligenz: Ein Internationales Handbuch* (S. 275-290). Göttingen: Hogrefe.
- Paulhus, D. L., Lysy, D. C., & Yik, M. S. (1998). Self-Report Measures of Intelligence: Are They Useful as Proxy IQ Tests? *Journal of Personality*, 66 (4), pp. 525-554.
- Pérez, J. C., Petrides, K. V., & Furnham, A. (2006). Die Messung von emotionaler Intelligenz als Trait. In R. Schulze, P. A. Freund, & R. D. Roberts (Hrsg.), *Emotionale Intelligenz. Ein Internationales Handbuch* (S. 191-211). Göttingen: Hogrefe.
- Petrides, K. V., & Furnham, A. (2003). Trait emotional intelligence: Behavioural validation in two studies of emotion recognition and reactivity to mood induction. *European Journal of Personality*, 17, pp. 39-75.
- Ployhart, R. E., & Ehrhart, M. G. (2003). Be Careful What You Ask For: Effects of Response Instructions on the Construct Validity and Reliability of Situational Judgment Tests. *International Journal of Selection and Assessment*, 11 (1), pp. 1-16.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. *Journal of Applied Psychology*, 88 (5), pp. 879-903.

- Richman-Hirsch, W. L., Olson-Buchanan, J. B., & Drasgow, F. (2000). Examining the Impact of Administration Medium on Examinee Perceptions and Attitudes. *Journal of Applied Psychology*, 85 (6), pp. 880-887.
- Rindermann, H. (2007). *Emotionale-Kompetenz-Fragebogen. Ein Verfahren zur Einschätzung emotionaler Kompetenzen und emotionaler Intelligenz aus Selbst- und Fremdsicht*. Göttingen: Hogrefe.
- Roberts, R. D. (2000). *Openness conscientiousness extraversion agreeableness neuroticism index condensed (OCEANIC) test manual*. Sydney, Australia: E-intelligence Testing Products.
- Roberts, R. D., Schulze, R., O'Brien, K., MacCann, C., Reid, J., & Maul, A. (2006). Exploring the Validity of the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) with Established Emotions Measures. *Emotion*, 6 (4), pp. 663-669.
- Roberts, R. D., Schulze, R., Zeidner, M., Matthews, G., Freund, P. A., & Kuhn, J.-T. (2006). Emotionale Intelligenz: Verstehen, Messen und Anwenden - Ein Resümee. In R. Schulze, P. A. Freund, & R. D. Roberts (Hrsg.), *Emotionale Intelligenz: Ein internationales Handbuch* (S. 313-341). Göttingen: Hogrefe.
- Roberts, R. D., Zeidner, M., & Matthews, G. (2001). Does emotional intelligence meet traditional standards for an intelligence? Some new data and conclusions. *Emotion*, 1 (3), pp. 196-231.
- Rode, J. C., Mooney, C. H., Arthaud-Day, M. L., Near, J. P., Rubin, R. S., Baldwin, T. T., & Bommer, W. H. (2008). An examination of the structural, discriminant, nomological, and incremental predictive validity of the MSCEIT V2.0. *Intelligence*, 36 (4), pp. 350-366.
- Roseman, I. J., & Evdokas, A. (2004). Appraisals cause experienced emotions: Experimental evidence. *Cognition and Emotion*, 18 (1), pp. 1-28.
- Rossen, E., & Kranzler, J. H. (2009). Incremental validity of the Mayer-Salovey-Caruso Emotional Intelligence Test Version 2.0 (MSCEIT) after controlling for personality and intelligence. *Journal of Research in Personality*, 43 (1), pp. 60-65.
- Rost, D. H. (2009). *Intelligenz. Fakten und Mythen*. Weinheim: Beltz Verlag.
- Sachs, L., & Hedderich, J. (2006). *Angewandte Statistik. Methodensammlung mit R*. Berlin: Springer.
- Salovey, P., & Mayer, J. D. (1990). Emotional Intelligence. *Imagination, Cognition, and Personality*, 9, pp. 185-211.
- Salovey, P., Mayer, J. D., Goldman, S. L., Turvey, C., & Palfai, T. P. (1995). Emotional attention, clarity, and repair: Exploring emotional intelligence using the Trait Meta-Mood Scale. In J. W. Pennebaker (Ed.), *Emotion, disclosure, and health* (pp. 125-154). Washington, DC: American Psychological Association.
- Schlegel, K. (2013). *Improving the measurement of emotion recognition ability*. Univ. Genève: Thèse de doctorat.

- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, pp. 262-274.
- Schmidt-Atzert, L. (2003). Das Mögliche machen. (Diskussionsbeitrag zu Schuler, 2002; Emotionale Intelligenz – ein irreführender und unnötiger Begriff). *Zeitschrift für Personalpsychologie*, 2, S. 24-26.
- Schmidt-Atzert, L., & Bühner, M. (2002). Entwicklung eines Leistungstests zur Emotionalen Intelligenz. *Vortrag auf dem 43. Kongress der Deutschen Gesellschaft für Psychologie*. Berlin.
- Schmidt-Atzert, L., & Hüppe, M. (1996). Emotionsskalen EMO 16. Ein Fragebogen zur Selbstbeschreibung des aktuellen emotionalen Gefühlszustandes. *Diagnostica*, 42, S. 242-267.
- Schneider, T. R., Lyons, J. B., & Khazon, S. (2013). Emotional intelligence and resilience. *Personality and Individual Differences*, 55, pp. 909–914.
- Schuler, H. (2002). Emotionale Intelligenz - ein irreführender und unnötiger Begriff (Diskussionsforum). *Zeitschrift für Personalpsychologie*, 3, S. 138-140.
- Schulte, M. J., Ree, M. J., & Carretta, T. R. (2004). Emotional intelligence: not much more than g and personality. *Personality and Individual Differences*, 37, pp. 1059–1068.
- Schulze, R., Freund, P. A., & Roberts, R. D. (2006). *Emotionale Intelligenz: Ein Internationales Handbuch*. Göttingen: Hogrefe.
- Schutte, N. S., Malouff, J. M., Hall, L. E., Haggerty, D. J., Cooper, J. T., Golden, C. J., & Dornheim, L. (1998). Development and validation of a measure of emotional intelligence. *Personality and Individual Differences*, 25 (2), pp. 167-177.
- Schutte, N. S., Malouff, J. M., Thorsteinsson, E. B., Bhullar, N., & Rooke, S. E. (2007). A meta-analytic investigation of the relationship between emotional intelligence and health. *Personality and Individual Differences*, 42, pp. 921-933.
- Schwaninger, A. (2005). Objekterkennung und Signaldetektion. In B. Kersten, & M. Groner (Hrsg.), *Praxisfelder der Wahrnehmungspsychologie* (S. 108-132). Bern: Huber.
- Sharma, S., Gangopadhyay, M., Austin, E., & Mandal, M. K. (2013). Development and Validation of a Situational Judgment Test of Emotional Intelligence. *International Journal of Selection and Assessment*, 21 (1), pp. 57-73.
- Siebert, K. (2007). *Zum Einfluss von Emotionaler Intelligenz auf die Übereinstimmung von Selbst- und Fremdeinschätzungen in der Persönlichkeitsforschung*. Digitale Dissertation, Universität Heidelberg.
- Siemer, M., Mauss, I., & Gross, J. J. (2007). Same Situation—Different Emotions: How Appraisals Shape Our Emotions. *Emotion*, 7 (3), pp. 592–600.

- Sjöberg, L. (2001). Emotional intelligence: A psychometric analysis. *European Psychologist*, 6, pp. 79-95.
- Spector, P. E. (2006). Method Variance in Organizational Research: Truth or Urban Legend? *Organizational Research Methods*, 9 (2), pp. 221-232.
- Spector, P. E., & Johnson, H.-A. M. (2009). Improving the Definition, Measurement, and Application of Emotional Intelligence. In K. R. Murphy (Ed.), *A Critique of Emotional Intelligence. What are the problems and how can they be fixed?* (pp. 325-344). New York: Routledge.
- Stankov, L. (1997). *Gf-Gc quickie test battery*. Sydney, Australia: School of Psychology.
- Steinmayr, R., Schütz, A., Hertel, J., & Schöder-Abé, M. (2011). Mayer-Salovey-Caruso Test zur Emotionalen Intelligenz (MSCEIT™). Deutschsprachige Adaptation des Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT™) von John D. Mayer, Peter Salovey und David R. Caruso. Manual. Bern: Verlag Hans Huber.
- Suzuki, A., Hoshino, T., & Shigemasu, K. (2010). Happiness is unique: A latent structure of emotion recognition traits revealed by statistical model comparison. *Personality and Individual Differences*, 48, pp. 196-201.
- Tapia, M. (2001). Measuring Emotional Intelligence. *Psychological Report*, 88, pp. 353-364.
- Thomson, F. (2009). *Überprüfung und Validierung eines videobasierten Testverfahrens zur Messung der Emotionalen Intelligenz von Mitarbeitern im Pflegebereich*. Unveröffentlichte Diplomarbeit. Marburg: Philipps-Universität.
- TIME. (1995, October 02). Emotional Intelligence: The EQ-Factor. *Time Warner, New-York*.
- van der Zee, K., Thijs, M., & Schakel, L. (2002). The Relationship of Emotional Intelligence with Academic Intelligence and the Big Five. *European Journal of Personality*, 16, pp. 103-125.
- Van Rooy, D. L., Viswesvaran, C., & Pluta, P. (2005). An evaluation of construct validity: What is this thing called Emotional Intelligence? *Human Performance*, 18, pp. 445-462.
- Visser, B. A., Ashton, M. C., & Vernon, P. A. (2008). What Makes You Think You're so Smart? Measured Abilities, Personality, and Sex Differences in Relation to Self-Estimates of Multiple Intelligences. *Journal of Individual Differences*, 29 (1), pp. 35-44.
- Visser, B. A., Bay, D., Cook, G. L., & Myburgh, J. (2010). Psychopathic and antisocial, but not emotionally intelligent. *Personality and Individual Differences*, 48, pp. 644-648.
- Watzlawick, P. (2004). *Anleitung zum Unglücklichsein*. München: Piper Verlag GmbH.
- Webb, C. A., Schwab, Z. J., Weber, M., DelDonno, S., Kipman, M., Weiner, M. R., & Killgore, W. D. (2013). Convergent and divergent validity of integrative versus mixed model measures of emotional intelligence. *Intelligence*, 41, pp. 149-156.
- Weekley, J. A., & Ployhart, R. E. (2006). *Situational Judgment Tests Theory, Measurement and Application*. San Francisco: Lawrence Erlbaum Associates.

- Weijters, B., Geuens, M., & Schillewaert, N. (2010). The Stability of Individual Response Styles. *Psychological Methods*, 15 (1), pp. 96-110.
- Wickens, T. D. (2002). *Elementary Signal Detection Theory*. New York: Oxford University Press.
- Wilhelm, O. (2005). Measures of emotional intelligence: practice and standards. In R. Schulze, & R. D. Roberts, *Emotional intelligence: An international handbook* (pp. 131-154). Göttingen: Hogrefe.
- Witten, T. (2003). *Emotionale Intelligenz im Berufsalltag: Entwicklung und Evaluierung eines Leistungstests*. Unveröffentlichte Diplomarbeit. Marburg: Phillips-Universität.
- Wonderlic Personnel Test Inc. (2000). *Wonderlic Personnel Test and scholastic level exam user's manual*. Libertyville, IL: Wonderlic Inc.
- Wong, C. S., & Law, K. S. (2002). The effects of leader and follower emotional intelligence on performance and attitude: An exploratory study. *The Leadership Quarterly*, 13, pp. 243-274.
- Zeidner, M., & Olnick-Shemesh, D. (2010). Emotional intelligence and subjective well-being revisited. *Personality and Individual Differences*, 48, pp. 431-435.
- Zeidner, M., Matthews, G., & Roberts, R. D. (2001). Slow Down, You Move Too Fast: Emotional Intelligence Remains an "Elusive" Intelligence. *Emotion*, 1, pp. 265-275.
- Zeidner, M., Matthews, G., Roberts, R. D., & MacCann, C. (2003). Development of Emotional Intelligence: Towards a Multi-Level Investment Model. *Human Development*, 46, pp. 69-96.
- Zeidner, M., Shani-Zinovich, I., Matthews, G., & Roberts, R. D. (2005). Assessing emotional intelligence in gifted and non-gifted high school students: Outcomes depend on the measure. *Intelligence*, 33, pp. 369-391.
- Zysberg, L., Levy, A., & Zisberg, A. (2011). Emotional Intelligence in Applicant Selection for Care-Related Academic Programs. *Journal of Psychoeducational Assessment*, 29 (1), pp. 27-38.

8 Erklärung

Hiermit versichere ich, Tonja Berit Plücken, dass ich die vorliegende Dissertation „Methodische Fragen bei der Erfassung von Emotionaler Intelligenz“ selbständig und ohne unerlaubte Hilfe verfasst, nicht andere als die in ihr angegebenen Quellen oder Hilfsmittel benutzt und alle vollständig oder sinngemäß übernommenen Zitate als solche gekennzeichnet habe.

Diese Dissertation wurde in der vorliegenden oder einer ähnlichen Form noch bei keiner anderen in- oder ausländischen Hochschule anlässlich eines Promotionsgesuchs oder zu anderen Prüfungszwecken eingereicht.

Stuttgart, den 26. Januar 2015

Tonja Berit Plücken



9 Lebenslauf Tonja Berit Plücken

Die Seite 173 (Lebenslauf) enthält persönliche Daten. Sie ist deshalb nicht Bestandteil der Online-Veröffentlichung.

